

Institut d'études politiques de Paris

ECOLE DOCTORALE DE SCIENCES PO

**Programme doctoral de science politique, mention sociologie politique et
action publique**

Centre d'études européennes de Sciences Po

Doctorat en science politique

Politiques des données urbaines

Ce que l'open data fait au gouvernement urbain

Antoine COURMONT

Thèse dirigée par Dominique BOULLIER, Professeur des Universités

Soutenue le 16 décembre 2016

Jury :

M. Dominique BOULLIER, Professeur des Universités en sociologie, Sciences Po Paris et École Polytechnique Fédérale de Lausanne, (directeur de thèse)

M. Geoffrey C. BOWKER, Professor, The Donald Bren School of Information and Computer Sciences, University of California (rapporteur)

M. Emmanuel DIDIER, chargé de recherche au CNRS, directeur adjoint d'Epidapo Lab, Institute for Society and Genetics, University of California

Mme Delphine GARDEY, professeure ordinaire, Université de Genève

M. Patrick LE GALÈS, directeur de recherche au CNRS, Centre d'études européennes, Sciences Po Paris

M. Gilles PINSON, Professeur des Universités en science politique, Sciences Po Bordeaux (rapporteur)

Remerciements

A l'instar de l'ouverture des données, la thèse est un processus qui se forge et s'affine au travers de différentes épreuves. Le contenu de ce manuscrit doit beaucoup aux différentes personnes rencontrées au cours de ces cinq années de thèse.

Mes remerciements sont adressés en premier lieu à Dominique Boullier qui a accepté de diriger cette thèse et qui m'a prodigué de précieux conseils tout au long de ces cinq années. Je souhaite lui exprimer mon immense gratitude pour son soutien, sa disponibilité et ses remarques qui ont été fondamentaux au développement de cette thèse, mais aussi à ma formation de chercheur.

Je tiens à remercier Geoffey C. Bowker et Gilles Pinson qui m'ont fait l'honneur d'être rapporteurs de cette thèse et Emmanuel Didier, Delphine Gardey et Patrick Le Galès d'avoir accepté de participer à mon jury de thèse.

Je tiens à exprimer ma reconnaissance à la communauté urbaine de Lyon qui a accepté de financer cette thèse par l'intermédiaire d'une convention CIFRE. Je remercie en premier lieu Philippe Vial pour son intérêt pour cette thèse, et surtout Grégory Bourbon pour son encadrement professionnel et la grande liberté qu'il m'a laissé dans la rédaction de cette thèse. Je remercie toute l'équipe du « SWO » de m'avoir fait découvrir le Grand Lyon et encouragé quotidiennement tout au long de ces trois années en leur compagnie. Merci également à la DPDP qui m'a fait confiance pour prolonger les réflexions entamées dans cette thèse pour penser la « ville de demain ». Merci enfin à l'ensemble des personnes rencontrées au cours de ce terrain lyonnais pour leur disponibilité et leur confiance, particulièrement Grégory et Laurence de « l'équipe projet » *open data*. Sans leur ouverture, ce travail de recherche ethnographique n'aurait pas été possible. Merci à Cécile, dont la Cifre sur la direction de l'eau de la Communauté urbaine m'a fait découvrir d'autres facettes de cette institution communautaire.

Merci aux chercheurs et à toute l'équipe administrative du Centre d'études européennes pour leur accueil, leur accompagnement et leurs conseils. Être intégré à une vie de laboratoire malgré mon éloignement lyonnais a été une grande aide pour la réalisation de cette thèse. Merci à la team des doctorants qui a rendu ce travail moins solitaire au travers des déjeuners,

des blagues et des échanges : Thomas, Francesca, Nuria, Antoine, Simon, Pauline, Camille, Vlad, Mickael, Clément et tous les autres. Merci particulièrement à Léa, Manon, Marine et Tom pour ces années et ces péripéties partagées. Merci à Nicolas, relecteur méticuleux et critique, et avec qui nous avons tenté de démêler les ficelles de l'ANT, de la Cifre, de la recherche et de son « monde ».

Le travail de terrain aux États-Unis a été soutenu par l'École doctorale de Sciences Po. Je remercie le département de sociologie de Columbia University de m'avoir accueilli durant ces trois mois à New-York et à la Pr. Saskia Sassen de m'avoir fourni de précieux conseils. Merci également à toutes les personnes dont les remarques, les questions et les critiques ont fait avancer mes réflexions et contribué à cette recherche.

J'ai une pensée particulière pour tous ceux qui ont accompagné quotidiennement, de près ou de loin, cette thèse. Martha, dont le soutien, la confiance et les encouragements ont été précieux pour mener à bien ce travail. À mes amis, qui se retrouveront dans cette thèse : rassurez-vous, vous êtes bien plus que les prénoms (anonymisés) de mes enquêtés. Un énorme merci à vous tous. Merci tout particulièrement à Mathias, Meryl, Ophélie, Etienne, Guillaume et Mathis pour leurs canapés à Lyon, Paris et New York. Merci enfin à mes parents, pour leur profonde affection et leur soutien indéfectible tout au long de ces (longues) années d'étude.

Résumé

Au travers de l'étude de la politique d'*open data*, cette thèse analyse ce que la mise en circulation des données fait au gouvernement urbain. En effectuant une sociologie des données attentive aux transformations conjointes des données et des acteurs qui leur sont associés, la thèse pointe le pluralisme des politiques des données urbaines entendues comme les modes de composition de collectifs autour des données.

La thèse repose sur une enquête ethnographique réalisée au sein de la Métropole de Lyon au cours de laquelle la politique d'ouverture de données de l'institution a été analysée *en train de se faire*. Ce travail ethnographique a été complété par la réalisation de 70 entretiens, le dépouillement d'archives communautaires et une comparaison partielle avec des villes nord-américaines.

En suivant la chaîne des données ouvertes, la thèse met en évidence une tension entre attachement et détachement. Attachées à de vastes réseaux sociotechniques dans lesquelles elles sont insérées, les données doivent être détachées de leur environnement initial pour être mise en circulation, avant d'être ré-attachées à de nouveaux utilisateurs. Pour cela, elles subissent une série d'épreuves, au résultat incertain, desquelles émergent de nouveaux agencements qui peuvent questionner les frontières, sectorielles, institutionnelles et territoriales du gouvernement urbain. Pour conserver la maîtrise de ses politiques publiques, l'enjeu pour une collectivité est dès lors de faire de la donnée un objet de gouvernement pour parvenir à réguler les flux de données sur son territoire.

Table des matières

Remerciements	5
Résumé	8
Table des matières	9
Liste des encadrés	15
Liste des illustrations	16
Liste des sigles	17
<i>Introduction</i>	19
I. La smart city : quand la donnée arrive en ville	20
II. Une sociologie des données	26
III. Politiques des données urbaines	30
IV. L’open data : la mise en circulation des données urbaines	32
V. Méthodologie	34
A. Le doctorant CIFRE : défis et opportunités du chercheur acteur de son objet	35
1) Du « Eux » au « Nous ». Entre observation et participation : travailler sur l’institution et pour l’institution	35
2) Rendre compte de l’enquête Cifre : le choix du récit ethnographique	38
B. L’étude d’un cas : la Métropole du Grand Lyon.....	39
C. Une étude centrée sur des données spatialisées	41
VI. Suivre la donnée : Les étapes du récit	42
<i>Chapitre 1 Produire : Attachements des données et agencements territoriaux</i>	45
I. Les données sont faites d’attachements	53
A. Du signal au centre de calcul : la chaîne de la « donnée » de circulation routière ..	54
B. Des attachements fragiles quotidiennement solidifiés	62
C. Conclusion : le réseau sociotechnique de la donnée	64
II. Représenter l’espace urbain : conventions et usages des données	65
A. Des données peu « réalistes » : représenter l’espace urbain pour agir dessus	65
1) La pluralité des représentations de l’espace urbain	66
2) Des représentations partielles de l’espace urbain	69
3) Les données reposent sur des conventions.....	70
B. Les effets politiques des conventions : transports et mobilité urbaine	72
C. Conclusion : les totalisations orientées de l’espace urbain	75
III. Les données comme outil de coordination : une représentation partagée du territoire métropolitain	78

A. Données de référence et données métiers : un investissement de forme d'ampleur différente.....	80
1) Données métiers et données de référence : une distinction par les usages ?	82
2) L'ampleur de « l'investissement de forme ».....	89
B. Le SIG et l'émergence d'un agencement territorial.....	91
1) Le SIG : un objet-frontière.....	91
2) Un agencement territorial	92
IV. Conclusion : données et territorialisation.....	95

Chapitre 2 Mettre en circulation : Généalogie de la politique d'ouverture des données de la communauté urbaine de Lyon 99

I. Trois arènes indépendantes de problématisation de la diffusion de données publiques..... 103

A. Arène 1 : La traduction communautaire de la directive européenne INSPIRE	104
1) La diffusion des données géographiques au Grand Lyon.....	105
2) La directive européenne INSPIRE impose aux administrations publiques la diffusion de leurs informations géographiques.....	105
3) La traduction de la directive INSPIRE par la communauté urbaine de Lyon ...	107
B. Arène 2 – La Centrale de mobilité : diffuser les données pour répondre aux enjeux environnementaux, économiques et de mobilité.....	109
1) La Centrale de Mobilité : la diffusion des données au cœur de la politique de mobilité	109
2) Le projet Optimod : la diffusion des données mobilité au service de l'innovation et du développement économique.....	111
C. Arène 3 – Le mouvement de l'open data	113
1) Les origines anglo-saxonnes du mouvement de l' <i>open data</i>	113
2) La réception de l'open data en France	117
3) Le Grand Lyon à la « mode » de l'open data ?.....	122

II. Le projet « ouverture des données publiques du Grand Lyon » : composer avec des intérêts hétérogènes..... 127

A. Explorer les possibles de l'ouverture des données	128
1) Rencontrer les producteurs de données.....	130
2) Rencontrer les réutilisateurs de données.....	132
B. Définir une stratégie : « Ce n'est pas un projet d'open data, c'est un projet d'ouverture des données »	134
1) La Smart City : nouveaux services et développement économique	135
2) Le poids d'Optimod : plateforme d'agglomération et respect des politiques publiques.....	139
(a) La plateforme du Grand Lyon ou la plateforme du territoire du Grand Lyon ?	139
(b) Laisser-faire ou contrôler les réutilisations ?	143
3) Une plateforme au service des politiques du Grand Lyon et de la ville intelligente	145
C. Faire valider la stratégie : les rencontres avec les élus	147

III. Elaborer les licences de réutilisation : une mise en boîte noire de la stratégie de diffusion	149
A. Associer des acteurs et des données hétérogènes en adoptant des licences génériques	151
B. ODbL versus Licence ouverte : une première licence pour favoriser les réutilisations commerciales.....	152
C. Une licence « avec authentification » pour s’assurer du respect de l’intérêt général	156
D. Une licence avec redevance pour éviter les monopoles.....	158
1) Identifier des critères objectifs de discrimination.....	159
2) Déterminer le montant de la redevance	161
IV. Conclusion – Pluralisme des politiques d’open data, pluralisme du rôle du grand lyon.....	164

Chapitre 3 Ouvrir : Les données à l’épreuve de leur diffusibilité..... 167

I. Identifier	172
A. Explorer l’espace des données de l’agglomération lyonnaise	172
1) La carte et le périmètre des « données candidates à l’ouverture ».....	174
2) Rencontrer les administrateurs de données géographiques	175
3) Lister les données d’un secteur d’action publique.....	177
4) Demander aux réutilisateurs	178
5) Multiplier les réunions exploratoires avec des acteurs divers	179
6) Regarder ce qui se fait « ailleurs ».....	180
II. Publiciser	183
A. Les « porte-paroles » de l’ouverture : quand les producteurs questionnent leurs données par rapport à des usages imaginés.....	184
B. Questionner la donnée : associer la donnée à un nouvel usage	187
1) Propriété et responsabilité.....	188
2) Qualité.....	191
3) Sensibilité : la figure de l’utilisateur malveillant.....	194
4) Intérêt : l’absence d’utilisateurs imaginés.....	198
5) La crainte d’un changement des relations de pouvoir	199
C. Enrôler les producteurs : les dispositifs d’intéressement.....	201
1) Réduire l’incertitude par la mise en avant d’exemples d’utilisation.....	202
2) Faciliter la diffusion des données en ayant recours à la plateforme open data..	204
3) Valoriser le travail des producteurs	206
4) Les licences comme dispositif d’intéressement.....	207
D. Qui décide de l’ouverture des données ?	208
III. Extraire	216
A. Extraire la donnée	217
1) L’extraction manuelle	217

2) L'extraction automatique	218
B. Travailler la donnée : le « nettoyage »	220
C. Diffuser la donnée : plateforme, formats et métadonnées comme « opérateurs de circulation »	228
1) L'infrastructure de diffusion	228
2) Le « formatage » : des formats ouverts pour favoriser l'usage large des données	229
3) Les métadonnées	233
D. La production des « données ouvertes » : un investissement de forme	235
IV. Conclusion – De la « donnée » à la « donnée ouverte »	237

Chapitre 4 Réutiliser : Le pluralisme des politiques de réutilisation des données ouvertes..... 241

I. Consolidation.....	246
A. Quelques situations de déliquescence des données ouvertes	248
1) Le manque d'informations contraint l'usage alternatif de la donnée	248
2) L'échelle est limitée au périmètre institutionnel du producteur	249
3) La structuration hétérogène des données	250
B. Opérations de consolidation et estimation de la « flexibilité interprétative » des données ouvertes	252
1) Le « sourcing » : identifier, comprendre et estimer la fiabilité de la donnée	253
2) Le nettoyage : préparer la donnée pour un usage secondaire	257
C. L'infrastructure partagée : propriété essentielle des données ouvertes en tant qu'objet-frontière	261
II. Homogénéisation	264
A. Des données ouvertes hétérogènes aux périmètres limités	265
B. Les opérations d'homogénéisation des données ouvertes	269
1) Identifier les données	269
2) Analyser le cadre juridique	271
3) Accéder aux données (et les actualiser)	272
4) Intégrer (et homogénéiser) les données	275
(a) La diversité des bases de données, l'impossible automatisation de l'homogénéisation	275
(b) OSM : le refus de l'import massif, la priorité accordée aux contributeurs	277
C. Google : une politique alternative d'homogénéisation des données publiques	280
D. De nouvelles conventions d'équivalence indépendantes des acteurs publics.....	284
1) Le GTFS : un standard imposant une nouvelle représentation des réseaux de transport en commun.....	284
2) L'effet de la hiérarchisation alternative des voies sur la politique de régulation du réseau routier.....	286
E. Conclusion : le modèle de la « plateforme » : une stratégie d'homogénéisation ..	288
III. Articulation.....	291

A.	Associer des jeux de données hétérogènes pour produire une nouvelle représentation de l'espace urbain.....	292
B.	Le travail d'articulation.....	294
1)	Les « opérateurs d'articulation » : la donnée-pivot	295
2)	Un travail d'articulation spécifique à chaque enjeu traité	301
C.	Nouvelle représentation de l'espace urbain, nouvel objet de gouvernement.....	304
1)	Du gouvernement des transports au gouvernement de la mobilité.....	305
(a)	L'utilisateur au centre d'un réseau de données.....	305
(b)	Un usager pluriel plutôt que des flux	308
D.	Conclusion : hétérogénéité des données et singularisation des entités représentées	309
IV.	Conclusion	310

Chapitre 5 Gouverner : Instaurer une politique de régulation des flux de données 313

I.	Formaliser la gouvernance de la donnée	317
A.	La gouvernance comme processus décisionnel de l'ouverture des données	318
B.	Administrer les données métropolitaines.....	321
C.	Une politique publique de la donnée métropolitaine ?	325
II.	Institutionnaliser la gouvernance de la donnée.....	328
A.	Le refus d'autonomiser la politique d'ouverture des données	328
B.	Des systèmes d'information au numérique : une autonomisation des services informatiques	329
C.	Du service de l'information géographique à l'administration des données métropolitaines.....	331
D.	Une nouvelle fonction : l'administratrice générale des données	333
III.	Instrumenter la gouvernance de la donnée	341
A.	Le catalogue : rendre gouvernables les données.....	343
B.	Licences & Conventions : devenir tiers de confiance.....	346
C.	Le label : gouverner par les modèles	350
D.	Le living lab : mettre en relation producteurs et utilisateurs	352
IV.	Légitimer la gouvernance de la donnée	358
A.	Une position de « tiers de confiance » : l'exemple de la mobilité.....	360
1)	Rassembler les données	361
2)	Réguler l'usage des données.....	362
3)	Acquérir une légitimité en jouant un rôle de tiers de confiance	365
B.	Des associations mises à l'épreuve, une légitimité contestée	367
1)	Des médiations fragiles.....	367
2)	Des producteurs non associés	368
3)	Des utilisateurs en dissidence	369
4)	Le « gouvernement à distance » des données métropolitaines par l'Union européenne et l'Etat	371

V. Conclusion	374
<i>Conclusion générale Ce que la mise en circulation des données fait au gouvernement urbain.....</i>	379
I. Des politiques plurielles d’open data.....	382
II. Ce que la mise en circulation des données fait au gouvernement urbain	383
A. La mise en circulation des données fait de la donnée un enjeu de gouvernement.	384
B. La mise en circulation des données questionne les frontières sectorielles, institutionnelles et territoriales.....	385
III. Les pistes de prolongement de la recherche : pour une sociologie politique des données urbaines.....	387
A. Nouvelles formes de quantification et nouveaux acteurs de gouvernement urbain	388
B. Le pluralisme des agencements territoriaux	390
<i>Sources</i>	393
IV. Bibliographie	394
V. Liste des entretiens.....	419
A. France.....	419
1) Grand Lyon	419
2) Producteurs et/ou utilisateurs de données	420
B. Etats-Unis.....	421
1) New York.....	421
2) Boston	422
3) Philadelphie.....	422

Liste des encadrés

Encadré 1 : Données, jeu de données, base de données	46
Encadré 2 : Les limites de la quantification.....	63
Encadré 3 : Open data & Open government.....	115
Encadré 4 : Les villes américaines s’emparent de l’open data. L’exemple de NYC.....	116
Encadré 5 : L’Union européenne : une logique de valorisation économique des données publiques	118
Encadré 6 : L’open data de l’État français : d’une approche patrimoniale aux données comme bien commun.....	118
Encadré 7 : Le Grand Lyon : seul acteur légitime pour constituer une plateforme territoriale ?.....	141
Encadré 8 : Les hackathons, des dispositifs de mobilisation des utilisateurs ou d’intéressement des producteurs ?.....	204
Encadré 9 : La mise à disposition des données du SYTRAL	212
Encadré 10 : L’impossible extraction des données.....	219
Encadré 11 : L’open data par défaut facilite-t-il la diffusibilité des données ?.....	239
Encadré 12 : Comment identifier les réutilisateurs des données ?.....	245
Encadré 13 : La production des cartes Google Maps.....	267
Encadré 14 : L’essor des data scientists, de nouveaux acteurs administratifs ?.....	300
Encadré 15 : La fonction de Chief Data Officer dans les administrations municipales américaines	335
Encadré 16 : Les élus et la donnée : le difficile passage du registre technique aux discours politiques	340
Encadré 17 : Le catalogue, un instrument au cœur des politiques de la donnée des villes américaines	345

Liste des illustrations

Illustration 1 : Pyramide des connaissances.....	48
Illustration 2 : Boucles à induction détectant le passage ou la présence d'un véhicule sur la chaussée.....	55
Illustration 3 : Plan du réseau de télécommunications de régulation des feux de circulation.....	55
Illustration 4 : Mesures d'une boucle de comptage (débit et taux à la minute).....	56
Illustration 5 : Mesures d'une boucle de comptage (évolution des débits et des taux sur une période de six minutes).....	57
Illustration 6 : Mesures d'une boucle (débit et vitesse moyenne) sur une plage de 24h.....	57
Illustration 7 : Opérateurs du PC Criter de la Métropole de Lyon.....	58
Illustration 8 : Plan de feux d'un carrefour (Pont Lafayette).....	59
Illustration 9 : « Ondes vertes » : programmation d'un passage au vert successif des feux du cours Lafayette.....	59
Illustration 10 : Contrôleur de carrefour à feux.....	60
Illustration 11 : Extraits des jeux de données « Emplacement moto de la ville de Lyon » et « Parc 2RM ».....	66
Illustration 12 : Capture d'écran du site d'information Onlymoov.com.....	70
Illustration 13 : Les « couches » du SIG articulées par un référentiel unique de localisation.....	81
Illustration 14 : Schéma de présentation du SIG et direction métiers correspondante ...	86
Illustration 15 : Document de présentation du projet open data du Grand Lyon.....	129
Illustration 16 : Document de présentation de la plateforme open data.....	147
Illustration 17 : Cartographie théorique du périmètre informationnel concerné par l'open data.....	175
Illustration 18 : Tableau de priorisation de l'ouverture des données.....	176
Illustration 19 : Tableau listant les données souhaitées pour l'intégration dans la Centrale de mobilité.....	178
Illustration 20 : Liste des données communales publiées sur les sites open data de collectivités.....	182
Illustration 21 : Document interne de présentation de la démarche de diffusion des données au comité de direction de la direction de la propreté (4 avril 2012).....	186
Illustration 22 : Document interne de présentation de la démarche de diffusion des données au comité de direction de la direction de la propreté (4 avril 2012).....	186
Illustration 23 : Document interne de présentation de la démarche de diffusion des données au comité de direction de la direction de la propreté (4 avril 2012).....	187
Illustration 24 : Flux des arrivées de l'aéroport Lyon Saint-Exupéry.....	224
Illustration 25 : La donnée « Toilettes publics » extraite de la base métier.....	226
Illustration 26 : La donnée « Toilettes publics » disponible sur SmartData.....	226
Illustration 27 : Extrait de la donnée « Marchés forains du Grand Lyon ».....	258
Illustration 28 : Avertissements de maintenance affichés sur le site data.grandlyon.com.....	275
Illustration 29 : Capture d'écran de l'application Optymod.....	307
Illustration 30 : Schéma de gouvernance de l'ouverture des données envisagé.....	320
Illustration 31 : Visuel de présentation des « contributeurs » du Tuba.....	355

Liste des sigles

AOT	Autorité organisatrice des transports
API	Application Programming Interface
BRGM	Bureau de recherches géologiques et minières
CCI	Chambre de commerce et d'industrie
CDO	Chief Data Officer
CG	Conseil général
CORALY	Coordination et Régulation du trafic de l'Agglomération Lyonnaise
Courly	Communauté urbaine de Lyon
CSV	Comma-separated values
DDE	Direction départementale de l'équipement
DDEES	Délégation au développement économique, emploi et savoirs
DE	Direction de l'eau
DG	Direction Générale
DGA	Directeur général adjoint
DGDEI	Délégation générale au développement économique et internationale
DGDU	Délégation générale au développement urbain
DGFIP	Direction générale des Finances publiques
DGI	Direction général des impôts
DGR	Délégation générale aux ressources
DINSI	Direction à l'innovation numérique et aux systèmes d'information
DP	Direction de la propreté
DPDP	Direction de la prospective et du dialogue public
DSIT	Direction des systèmes d'information et des télécommunications
DSP	Délégation de Service Public
DV	Direction de la voirie
ERDF	Électricité Réseau Distribution France
GDF	Gaz de France
GTFS	General Transit Feed Specification
IDG	Infrastructure de données géographiques
IGN	Institut national de l'information géographique et forestière
INSEE	Institut national de la statistique et des études économiques
INSPIRE	Infrastructure for Spatial Information in the European Community
IRIS	Ilôts regroupés pour l'information statistique
LPA	Lyon Parc Auto
MOI	Management Organisation Informatique
OdBL	Open Database Licence
OGC	Open Geospatial Consortium
PDU	Plan de Déplacement Urbain
PLU	Plan local d'urbanisme
POI	Point of Interests
RDS-TMC	Radio System Data / Traffic Message Channel
RIL	Répertoire des Immeubles Localisés
SDIS	Service départemental d'incendie et de secours

SI	Système d'information
SIG	Système d'information géographique
SUR	Système urbain de référence
SWO	Service Web et Organisation
SYTRAL	Syndicat des Transports de l'Agglomération Lyonnaise
TCL	Transport en Commun Lyonnais
UE	Union européenne
VLS	Vélo en libre service
WFS	Web Feature Service
WMS	Web Map Service

Introduction

« *In God we trust, everyone else, bring data* »

Michael Bloomberg, ancien maire de New York, a fait de cette citation, attribuée au statisticien W. Edwards Deming, sa devise. Fidèle à cette croyance messianique en la donnée, le fondateur de Bloomberg LP, une société de services et d'informations financières, a encouragé l'usage des données au sein de l'administration durant ses trois mandats (2002-2013). « *I've long said that my time in business and government has taught me that if you can't measure it, you can't manage it. Our Administration has made usage of data a hallmark of our problem-solving strategies*¹. » Ce recours massif à la donnée se manifeste par la mise en place de plusieurs initiatives : stratégie digitale, publication d'indicateurs de performance de l'activité des services, politique d'*open data*, mise en place d'un service de *data analytics*, etc.

La place croissante prise par les données dans le gouvernement des villes n'est pas réservée à l'administration new-yorkaise. Alors que l'on évoquait auparavant les statistiques ou les indicateurs, les années 2000 marquent l'apparition de nouvelles entités dans le gouvernement urbain : les données. Dans l'ensemble des secteurs d'action publique, ces entités « *font irruption de façon surprenante, en allongeant la liste de ceux qu'il faut prendre en compte*² ». Si cette « nouveauté » est à relativiser, la prolifération des données interroge les modes de gouvernement de la ville. En quoi l'émergence d'un « monde de données » recompose le gouvernement urbain ? Comment compose-t-on avec ces nouvelles entités ?

I. La smart city : quand la donnée arrive en ville

L'administration municipale a toujours été productrice et utilisatrice de savoirs de gouvernement. La construction du pouvoir urbain s'accompagne d'un ensemble de

¹ Source : *Expanding the use of data analytics in City Governments*, 9 novembre 2011, <https://www.mikebloomberg.com/news/expanding-the-use-of-data-analytics-in-city-governments> (Consulté le 9 mars 2016)

A l'issue de son mandat, Michael Bloomberg poursuit sa volonté d'encourager l'utilisation de données pour conduire les politiques urbaines au travers du programme *What work cities* de sa fondation Bloomberg Philanthropies.

Source : <http://whatworkscities.bloomberg.org> (Consulté le 9 mars 2016).

² Bruno Latour, *Politiques de la nature. Comment faire entrer les sciences en démocratie*, Paris, La Découverte, 1999, p. 120.

techniques, de savoirs et de connaissances nécessaires à son organisation et son fonctionnement, ainsi que de description du territoire, de sa population et de son économie. Avec l'apparition de l'informatique au sein des administrations locales dans les années soixante-dix¹, ces savoirs de gouvernement commencent à être numérisés, les registres deviennent des systèmes d'information, dont on découvre les capacités inédites de traitement et de calcul. Toutefois, si ces fichiers numériques sont composés de « données », ce dernier terme est peu usité. Il faut attendre la fin des années 2000 pour que les données émergent dans les discours et s'imposent comme une entité essentielle du gouvernement urbain. On entre alors dans l'ère du numérique, où l'informatique n'est plus localisable dans des machines mais devient *pervasive*².

L'apparition des données dans le gouvernement urbain³ est liée à la dissémination croissante des technologies de l'information et de la communication dans l'espace urbain. La fiction de l'informatique ubiquitaire, cet idéal de l'informatique omniprésente et invisible, est en partie réalisée grâce à la miniaturisation des capteurs, la diminution des coûts de stockage, l'augmentation considérable de la puissance de calcul et l'expansion des réseaux de communication⁴. Conséquence de cette prolifération de l'informatique dans l'espace urbain,

¹ Il convient de rappeler que la société informationnelle est bien antérieure à l'apparition de l'ordinateur. Comme l'a mis en avant Delphine Gardey, tout un ensemble de savoirs, de dispositifs techniques et de pratiques a précédé l'apparition de l'informatique, reconfigurant les possibilités d'interventions sur le monde et de gouvernement des populations. Delphine Gardey, *Ecrire, calculer, classer : Comment une révolution de papier a transformé les sociétés contemporaines*, Paris, La Découverte, 2008, 320 p.

² Ce terme anglais difficilement traduisible caractérise l'omniprésence de l'informatique dans notre environnement. Cf. Dominique Boullier, *Sociologie du numérique*, Paris, Armand Colin, 2016, p. 49.

³ Dans cette thèse, le terme de gouvernement urbain est employé pour désigner les modalités d'élaboration et de mise en oeuvre de politiques publiques par les acteurs de la ville. Il s'appuie sur le postulat, mis en avant par les travaux de sociologie politique urbaine, que les villes ne sont plus uniquement des espaces de mises en oeuvre de politiques publiques définies par ailleurs (dans une logique de rapports centre-périphérie à laquelle se sont longtemps cantonnés les travaux sur les politiques urbaines), mais qu'elles sont devenues des acteurs autonomes élaborant ces politiques, définissant des stratégies et produisant du sens.

Prenant acte de la complexification et de la pluralisation des politiques urbaines, le terme de gouvernement ne réduit pas les acteurs gouvernementaux aux institutions publiques agissant dans une logique verticale, mais à l'ensemble des acteurs, publics ou privés, qui participent à l'action publique urbaine. Au travers du terme de gouvernement urbain, l'objectif est de réfléchir aux modalités de coordination dans l'action collective à l'échelle de la ville.

Sur ces évolutions du gouvernement des villes, voir notamment : Patrick Le Galès, « Du gouvernement des villes à la gouvernance urbaine », *Revue française de science politique*, 1995, vol. 45, n° 1, p. 57-95 ; Dominique Lorrain et Patrick Le Galès, « Gouverner les très grandes métropoles », *Revue française d'administration publique*, 2003, vol. 3, n° 107, p. 305-317 ; Gilles Pinson, « La gouvernance des villes françaises. Du schéma centre-périphérie aux régimes urbains », *Pôle Sud*, 2010, vol. 1, n° 32, p. 73-92 ; Patrick Le Galès, *Le retour des villes européennes : Sociétés urbaines, mondialisation, gouvernement et gouvernance*, 2e édition., Paris, Presses de Sciences Po, 2011.

⁴ Antoine Picon, *Smart Cities. Théorie et critique d'un idéal auto-réalisateur.*, Paris, Editions B2, 2013, 117 p.

la diversité des sources de données s'accroît. Les collectivités multiplient les capteurs dans l'espace urbain pour détecter, mesurer et enregistrer ce qui se passe dans la ville. Les entreprises enregistrent une multitude de données transactionnelles. Dotés de téléphone portable, les individus produisent des traces numériques qui sont autant d'informations sur leurs comportements, conservées précieusement par des entreprises de l'économie numérique. Le numérique conduit à l'émergence des données dans le gouvernement urbain. Alors qu'elles étaient auparavant soigneusement conservées dans des silos structurés, enfouies au sein de systèmes d'information, les données pénètrent maintenant dans tout notre environnement. Elles ne font plus système avec l'informatique et les réseaux, elles se sont détachées, prolifèrent et circulent jusqu'à constituer une nouvelle « digital skin¹ » enveloppant la ville.

Cette « couche informationnelle » caractérise la *smart city*. Popularisé par l'entreprise IBM à des fins commerciales², la ville intelligente est une notion floue aux définitions multiples³, selon que l'accent soit mis sur les technologies, les acteurs ou le gouvernement⁴. On peut toutefois la définir de manière générique comme l'usage généralisé de données numériques pour rationaliser la planification et la gestion des villes⁵.

Les discours sur la ville intelligente présentent la ville comme une totalité fonctionnelle et cohérente. La smart city est en effet pensée dans une vision systémique du fonctionnement urbain : la plasticité de la matière informationnelle doit permettre de gérer les interdépendances entre les multiples fonctions urbaines. La « couche informationnelle » relie – et dépasse- les différents secteurs d'action publique formant une « digital skin » enrobant les villes⁶. Pour cela, il est nécessaire de décloisonner les « silos » de données et passer d'une

¹ Chirag Rabari et Michael Storper, « The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 27-42.

² Ola Söderström, Till Paasche et Francisco Klauser, « Smart cities as corporate storytelling », *City*, 2014, vol. 18, n° 3, p. 307-320.

³ Emmanouil Tranos et Drew Gertner, « Smart networked cities? », *Innovation: The European Journal of Social Science Research*, 2012, vol. 25, n° 2, p. 175-190.

⁴ Albert Meijer et Manuel Pedro Rodriguez Bolivar, « Governing the smart city: a review of the literature on smart urban governance », *International Review of Administrative Sciences*, 2016, vol. 82, n° 2, p. 392-408.

⁵ Anthony Townsend, *Smart Cities - Big Data, Civic Hackers, and the Quest for a New Utopia*, New York, W. W. Norton & Company, 2013, 400 p.

⁶ C. Rabari et M. Storper, « The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data », art cit.

logique de stock à une logique de flux. Cette vision traduit l'application du « libéralisme informationnel » à l'échelle des villes. Théorisé par Benjamin Loveluck à partir d'une généalogie politique d'Internet¹, le libéralisme informationnel est une reformulation de la théorie politique libérale fondée sur les flux d'informations au sein des réseaux informatiques. Le paradigme de la mobilité, mis en avant par John Urry², s'applique maintenant aux flux d'informations : les données doivent être mobiles. La *smart city* caractérise ainsi la mise en circulation des données à l'échelle des villes.

Cette injonction à la mise en circulation des données amplifie les caractéristiques d'accessibilité et de centralité attribuées aux villes³. L'accès facilité aux données doit permettre de mieux construire, gérer, réguler et gouverner la ville. Le fantasme d'une ville ubiquitaire et l'utopie cybernétique seraient à portée de réalisation. On est en mesure de connaître et d'agir en temps réel, à distance, sur l'espace urbain. Cela permet de renforcer la sécurité, l'environnement, la gestion des flux, etc. On peut également utiliser toutes ces données pour construire des modèles prédictifs, et ainsi rationaliser l'urbanisme et le développement urbain. En dotant les réseaux énergétiques de capteurs de données, les *smart grids* visent par exemple à provisionner au mieux la fourniture énergétique et optimiser la consommation des ressources rares. Des données plus complètes, plus accessibles et plus interconnectées doivent ainsi renforcer la rationalité de l'action gouvernementale et ainsi améliorer l'efficacité, l'équité et la qualité de vie des citoyens⁴. Outre les *outcomes* des politiques urbaines, la *smart city* vise aussi à améliorer ses *process*. Les TIC doivent permettre une organisation plus horizontale des relations de pouvoir, une plus grande transparence et des pratiques collaboratives. La réduction de l'asymétrie d'informations encourage l'*empowerment* des citoyens, améliore l'*accountability* et la conduite des affaires gouvernementales.

¹ Benjamin Loveluck, *Réseaux, libertés et contrôle. Une généalogie politique d'internet*, Paris, Armand Colin, 2015.

² John Urry, *Sociology beyond Societies. Mobilities for the twenty-First Century*, London, Routledge, 2000, 268 p.

³ Henri Lefebvre, *Le droit à la ville*, 3e édition., Paris, Economica, 2009, 135 p.

⁴ Michael Batty et al., « Smart Cities of the Future », *European Physical Journal Special Topics*, 2012, n° 214, p. 481-518.

Des politiques de contrôle des accès¹ aux données sont toutefois mises en place parallèlement à l'émergence de ces flux de données contribuant au renouvellement des centralités urbaines. En effet, pour certains auteurs critiques, les *smart cities* ne seraient que le nouvel avatar d'arguments de marketing urbain visant à masquer les recompositions de pouvoir sous la figure d'une technologie neutre². La smart city serait une ville dépolitisée : les enjeux politiques deviennent des problèmes techniques sous l'effet de l'idéologie du « solutionnisme » technologique³. Le risque est alors de voir se développer une gouvernance technocratique⁴. La centralisation des données renforcerait le pilotage de la ville par des experts. Elle présenterait un risque dystopique de surveillance généralisée, ou tout du moins de contrôle accru sur les individus⁵, qui remettrait en cause une des caractéristiques majeures de la ville : l'anonymat qu'elle permet⁶.

Ce recours massif à la technique et cette évolution vers un gouvernement d'experts contribueraient à accélérer le transfert de pouvoir des acteurs publics aux acteurs privés. La smart city ne serait alors que le nouvel avatar de la ville néolibérale qui se développe depuis les années 80, et qui a pris le nom successif de « competitive cities », « creative cities », « sustainable cities », « green cities », etc. La « smart city » est dès lors un concept idéologique qui cache une gouvernance entrepreneuriale des villes dans laquelle les entreprises privées jouent un rôle central⁷. Ce transfert de pouvoir vers des acteurs privés se manifeste par l'entrée de nouveaux acteurs dans la fabrique de la ville. Spécialistes des technologies de l'information et du traitement des données, des firmes telles que Cisco, IBM,

¹ Jeremy Rifkin, *L'âge de l'accès*, Paris, La Découverte, 2005, 406 p.

² Robert G. Hollands, « Will the real smart city please stand up? », *City*, 2008, vol. 12, n° 3, p. 303-320 ; O. Söderström, T. Paasche et F. Klauser, « Smart cities as corporate storytelling », art cit ; Donald McNeill, « Global firms and smart technologies: IBM and the reduction of cities », *Transactions of the Institute of British Geographers*, 2015, vol. 40, n° 4, p. 562-574.

³ Evgeny Morozov, *To Save Everything, Click Here: The Folly of Technological Solutionism*, New York, Public Affairs, 2013, 432 p.

⁴ Dan Hill, « On the smart city; Or, a “manifesto” for smart citizens instead », *City of Sound*, 1 février 2013 ; A. Townsend, *Smart Cities - Big Data, Civic Hackers, and the Quest for a New Utopia*, op. cit. ; Adam Greenfield, *Against the Smart City*, London, Do Projects, 2013.

⁵ Alberto Vanolo, « Smartmentality: The Smart City as Disciplinary Strategy », *Urban studies*, 2014, vol. 51, n° 5, p. 883-898 ; Francisco Klauser, Till Paasche et Ola Söderström, « Michel Foucault and the Smart City: Power Dynamics Inherent in Contemporary Governing through Code », *Environment and Planning D: Society and Space*, 2014, vol. 32, p. 869-885.

⁶ Georg Simmel, *Les grandes villes et la vie de l'esprit*, Paris, Payot, 2013, 112 p.

⁷ Robert G. Hollands, « Critical interventions into the corporate smart city », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 61-77 ; Alan Wiig, « IBM's smart city as techno-utopian policy mobility », *City*, 2015, vol. 19, n° 2-3, p. 258-273.

Siemens ou Toshiba voient dans les politiques urbaines un relai de croissance et entendent jouer un nouveau rôle dans la formulation et la mise en œuvre des services urbains¹.

Ces discours sur la ville intelligente mettent en avant la pluralité des modèles de villes intelligentes selon que l'accent soit mis sur les recompositions de l'accessibilité ou de la centralité liées à la mise en circulation des données. Dominique Boullier propose une typologie alternative des *smart cities* basée sur leur architecture technique. Le premier modèle est celui de la « Good old city » : les données sont conservées dans des « silos » aux mains des autorités traditionnelles. Le second, qu'il qualifie de moderniste, est « IBM City » : la modélisation systémique des villes par des experts techniques au travers d'une centralisation des données mise en scène dans des *dashboards* ou des centres de commandement, comme par exemple celui mis en place par l'entreprise IBM à Rio de Janeiro². Le troisième modèle est la « Google City » : l'agrégation et la corrélation de traces et de données sans modèle préétabli. Enfin, le dernier modèle est la « Wiki City » reposant sur les contributions des citoyens habitants qui peuvent mettre en avant leur expérience de la ville³.

*
* *

Les travaux existants sur les *smart cities* restent souvent cantonnés à l'analyse des discours des firmes et des acteurs gouvernementaux. Ces analyses ont tendance à figer les discours et à dériver vers un déterminisme technologique qui assigne une direction univoque à l'émergence des données dans le gouvernement urbain. Or, à l'instar du pluralisme des modèles de villes intelligentes pointé par Dominique Boullier, les quelques études empiriques révèlent que la *smart city* est loin de se résumer à un modèle uniforme mais varie entre et au sein des villes⁴.

¹ A. Greenfield, *Against the Smart City*, *op. cit.* ; D. McNeill, « Global firms and smart technologies: IBM and the reduction of cities », art cit.

² D. McNeill, « Global firms and smart technologies: IBM and the reduction of cities », art cit ; A. Townsend, *Smart Cities - Big Data, Civic Hackers, and the Quest for a New Utopia*, *op. cit.*, p. 57-92 ; Christopher Gaffney et Cerianne Robertson, « Smarter than Smart: Rio de Janeiro's Flawed Emergence as a Smart City », *Journal of Urban Technology*, 2016, p. 1-18.

³ D. Boullier, *Sociologie du numérique*, *op. cit.*, p. 243-246.

⁴ Comme le soulignent Shelton et al. « *not all places of the city will be equally smart* ».

Surtout, l'impensé de la littérature sur la ville intelligente est son absence de questionnement sur la donnée. Celle-ci est considérée comme neutre et disponible dans une perspective purement technique ou fonctionnaliste : la donnée existe, est autonome, détachée, et circule aisément. Or, lorsque l'on confronte l'analyse des discours aux pratiques quotidiennes, le constat est loin d'être aussi évident. L'analyse des processus d'ouverture des données révèle que les données ne sont pas des entités autonomes qui peuvent aisément être mises en circulation. Pour comprendre ce décalage, il est nécessaire d'effectuer une sociologie qui questionne conjointement la technique et le social en décrivant les agencements hybrides d'humains et de non-humains que les données forment et transforment lorsqu'elles sont mises en circulation.

II. Une sociologie des données

La question qui traverse cette thèse est la suivante : quelles sont les politiques de données urbaines qui émergent de la mise en circulation des données ? Pour y répondre, j'effectue une sociologie des données en les considérant comme des dispositifs sociotechniques qui recomposent les modes de gouvernement de la ville.

Plutôt que d'affirmer ce que les données sont, devraient être ou devraient faire, la perspective suivie dans cette thèse est de rendre compte des données *en train de se faire* et *en train d'agir*. Cette démarche s'inspire de la sociologie des sciences et des techniques. Ce courant de recherche, qui marque une rupture considérable par rapport aux travaux de philosophie des sciences d'après guerre¹, analyse la science *en train de se faire* en couplant une analyse des énoncés et des pratiques scientifiques. Il s'agit ainsi de dépasser l'opposition entre « internalisme » et « externalisme ». Dans le premier cas, l'étude se concentre sur les représentations contenues dans les données ; les aspects cognitifs sont autonomisés et étudiés pour eux-mêmes. Dans le second cas, l'analyse porte en priorité sur l'organisation sociale des activités de production et d'usage des données, mais ne rend pas compte de manière approfondie de leurs contenus. Les sociologues et les historiens des sciences ont contesté cette

Ugo Rossi, « The Variegated Economics and the Potential Politics of the Smart City », *Territory, Politics, Governance*, 2016, vol. 4, n° 3, p. 337-353 ; Taylor Shelton, Matthew Zook et Alan Wiig, « The “Actually Existing Smart City” », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 13-25.

¹ Dominique Pestre, *Introduction aux Science Studies*, Repères., Paris, La Découverte, 2006, 128 p.

opposition entre idées et pratiques sociales pour mettre en avant leur imbrication dans des supports matériels, des outils ou des inscriptions¹. Ancrés dans des réseaux sociotechniques, les savoirs ne peuvent être pensés de façon autonome.

La sociologie de l'acteur réseau (ANT) invite à étudier un objet en analysant la multiplicité des relations qui le constituent, leur évolution et la nature de ces liens afin de mettre en évidence l'*agency* de cet objet. Ces relations s'établissent par des opérations de « traduction » au cours desquels des acteurs se constituent en porte-parole, traduisent la volonté de collectifs et tentent d'enrôler de nouveaux acteurs. Cette perspective pragmatiste considère que toute personne, tout objet, ou même la société, est toujours le résultat provisoire d'un réseau. Elle ne préexiste pas, mais elle est le résultat d'un travail de traduction qui donne à chaque entité un rôle particulier. A ce titre, les données peuvent être considérées comme un dispositif sociotechnique, c'est-à-dire un « *assemblage d'actants humains et non humains où les compétences et les performances sont distribuées* »². L'enjeu d'une sociologie de la donnée est d'ouvrir les boîtes noires de ces instruments quantitatifs, d'en déployer le réseau qui les constitue, en étant attentif à l'ensemble des associations entre humains et non-humains étudiés symétriquement. Cela permet de mettre en évidence l'*agency* des données, c'est-à-dire le fait que les données *font* des choses comme l'ont pointé les travaux de sociologie de la quantification.

Inspiré par la théorie de l'acteur-réseau et l'économie des conventions, Alain Desrosières a développé en France la sociologie de la quantification, c'est-à-dire l'étude des processus de mise en nombre de ce qui était exprimé par des mots. Parvenant à maintenir conjointement une analyse « internaliste » et « externaliste » de la statistique, il analyse parallèlement les outils et les modèles statistiques et leur usage administratif et étatique. Par ce biais, Desrosières souligne le fait que la statistique *fait* des choses : la quantification « *ne fournit pas seulement un reflet du monde, mais elle le transforme, en le reconfigurant autrement* »³. Ted Porter pointe également ce « pouvoir créateur » de la quantification⁴, tout

¹ Bruno Latour, *La science en action : Introduction à la sociologie des sciences*, 3e éd., Paris, La Découverte, 2005.

² Madeleine Akrich et Bruno Latour, « A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies » dans Wiebe E. Bijker et John Law (eds.), *Haping Technology/Building Society. Studies in Sociotechnical Change*, Cambridge Mass., MIT Press, 1992, p. 259.

³ Alain Desrosières, *Prouver et gouverner. Une analyse politique des statistiques publiques.*, Paris, La Découverte, 2014, p. 39.

⁴ Theodore Porter, « Making things quantitative », *Science in context*, 1994, vol. 7, n° 3, p. 389-407.

comme Wendy Espeland et Mitchell Stevens qui soulignent la capacité de la quantification « *to remake what it measures* »¹. Emmanuel Didier met en évidence le pouvoir de la statistique de donner « consistance » à des agrégats sociaux qui peuvent être aussi important que l'Amérique². De manière similaire, la sociologie de la comptabilité, qui a émergé autour notamment des travaux d'Eve Chiapello en France et d'Anthony Hopwood, de Peter Miller et de la revue *Accounting, Organization and Society*³ en Grande-Bretagne, a mis en avant les évolutions de cet instrument et sa capacité à transformer ce qu'elle décrit⁴, participant ainsi à la construction de l'économie⁵. Principalement centrés sur la statistique et la comptabilité, ces travaux de sociologie de la quantification se sont peu intéressés à la question des *data*. Est-ce que les données, à l'instar de ces dispositifs de calcul, disposent d'une capacité à « changer le monde » en qualifiant des objets en vue de les faire advenir ? Quelle est l'*agency* des données ? Est-ce qu'elles « *font des choses. [elles] coordonnent les actions ; [elles] agissent ou font agir les autres* »⁶ ?

En science politique, la sociologie des sciences et des techniques a inspiré Pierre Lascoumes et Patrick Le Galès dans leur proposition de développer une approche par les instruments d'action publique⁷. Refusant toute perspective fonctionnaliste selon laquelle l'instrument ne serait qu'un outil neutre visant la résolution d'un problème spécifique, ils affirment que les IAP sont « *porteur d'une conception concrète du rapport politique/société*

¹ Wendy Nelson Espeland et Mitchell L. Stevens, « A Sociology of Quantification », *European Journal of Sociology*, 2008, vol. 49, n° 03, p. 431.

² Emmanuel Didier, *En quoi consiste l'Amérique ? Les statistiques, le New Deal et la démocratie*, Paris, La Découverte, 2009, 320 p.

³ Pour une étude de cette revue, voir Fabrice Bardet, *La contre-révolution comptable. Ces chiffres qui (nous) gouvernent*, Paris, Les Belles Lettres, 2014, 374 p.

⁴ Anthony Hopwood et Peter Miller (eds.), *Accounting as Social and Institutional Practice*, Cambridge, Cambridge University Press, 1994, 340 p.

⁵ Eve Chiapello, « Accounting and the Birth of the Notion of Capitalism », *Critical Perspective on Accounting*, 2007, n° 18, p. 263-296.

⁶ Michel Callon, Yuval Millo et Fabian Muniesa, *Market Devices*, Malden, Blackwell Publishing, 2007, p. 2.

⁷ L'étude des instruments d'action publique avait déjà été abordée en science politique, notamment autour des travaux de Christopher Hood, mais elle se focalisait principalement sur les critères de choix de ces instruments. Les intérêts et les rapports de force pour imposer l'instrument jugé le plus pertinent (ou le moins contraignant) étaient jugés comme les déterminants principaux de ce choix. Christopher Hood, *The Tools of Government*, London, Macmillan, 1983, 178 p ;

Sur la distinction entre ces deux approches des instruments d'action publique, voir Pierre Lascoumes et Louis Simard, « L'action publique au prisme de ses instruments. », *Revue française de science politique*, 2011, vol. 61, n° 1, p. 5-22.

*et soutenu par une conception de la régulation*¹ ». Cette approche par les instruments s'inscrit dans une perspective foucauldienne d'étude des dispositifs de gouvernementalité. Les données répondent à cette définition, Michel Foucault fait d'ailleurs fréquemment référence à la statistique comme mise en forme de la population visant à la gouverner².

Si elle reprend l'invitation de ces auteurs à être attentif à la « théorisation politique implicite³ », aux représentations du monde qu'il incorpore et aux effets propres de l'instrument, la sociologie des données que je développe dans cette thèse s'en écarte sur un point majeur. Lascoumes et Le Galès invitent à révéler les recompositions de l'État par l'analyse des instruments. Cette proposition repose sur le postulat que l'État est *déjà-là*, qu'il préexiste et qu'il est doté d'une autonomie⁴. Or, la sociologie de l'acteur-réseau met en avant le caractère dynamique et instable de composition des collectifs et des agrégats sociaux. Dans cette perspective, l'État, ou la société, sont toujours en train de se faire. Par sa capacité à assembler des éléments hétérogènes, la donnée peut certes être considérée comme un instrument, mais elle est avant tout un agencement politique qui fait exister un collectif. Tout l'enjeu de la sociologie de la donnée est d'être attentif aux associations, c'est-à-dire d'étudier ce qui fait tenir les choses ensemble et la manière dont le social se trouve continûment réassemblé.

La thèse effectue une sociologie des données entendue comme l'étude des assemblages avec et autour des données. Elle part des données pour étudier les réseaux spécifiques d'actants qui lui sont associés. Chaque détail, chaque spécificité, chaque traduction de la donnée modifie la composition possible du réseau qui se déploie autour de la donnée. Chacune de ces compositions constitue une politique des données urbaines.

¹ Pierre Lascoumes, « La Gouvernémentalité : de la critique de l'État aux technologies du pouvoir », *Le Portique*, 2004, n° 13-14, p. 7.

² Notamment dans ses cours au Collège de France de 1978. Michel Foucault, *Sécurité territoire population*, Paris, Seuil, 2004, 448 p.

³ P. Lascoumes, « La Gouvernémentalité : de la critique de l'État aux technologies du pouvoir », art cit, p. 9.

⁴ Sur la distinction entre sociologie de l'acteur-réseau et analyse de l'instrumentation de l'action publique, voir Liz McFall, « What Have Market Devices Got to Do with Public Policy? For Regulation, Market Devices and Policy Instruments Collection » dans Charlotte Halpern, Pierre Lascoumes et Patrick Le Galès (eds.), *L'instrumentation de l'action publique.*, Paris, Presses de Sciences Po, 2014, p. 119-142.

III. Politiques des données urbaines

L'approche de l'ANT invite à penser l'*agency* dont disposent les données pour inventer d'autres arrangements, pour assembler autrement des êtres. Mise en circulation, la donnée est l'objet d'appropriations et de transformations diverses, ce qui suscite la recomposition de plusieurs collectifs, brouillant les limites entre acteur public et acteur privé, technique et social, secteurs d'action publique, transparence et développement économique, etc.

Dans cette perspective, la thèse décrit les déploiements d'associations autour et avec les données, visant à unifier ce qui était dispersé, à rendre commun ce qui n'avait rien de commun. Dans la continuité de Bruno Latour¹, ces opérations de composition de collectifs autour des données peuvent être entendues comme des *politiques*. Ce travail de rassemblement, élément après élément, d'intérêts, de valeurs et d'entités incommensurables dans un *modus vivendi* commun, est toujours fragile et révisable. Une des spécificités de l'ANT est en effet de ne pas considérer le monde social comme figé, mais d'être attentif aux multiples épreuves subies par les entités, à l'ensemble des traductions successives, nécessaires à l'alignement d'acteurs aux intérêts pluriels et hétérogènes. Une hypothèse centrale guide mon approche : plusieurs modes de composition avec et autour des données coexistent. Ce sont ces *politiques* de la donnée que j'explore dans cette thèse.

Cette conception du *politique* va à l'encontre de la définition défendue par la science politique qui restreint le politique au champ politique. Comme le rappelle Pierre Favre dans sa controverse avec Bruno Latour, dans la conception de ce dernier tout serait politique. Or, pour Favre, un objet « *ne devient « politique » que si une ou plusieurs conditions supplémentaires sont remplies, l'analyse de ces conditions étant en elle-même un des objets centraux de la science politique*² ». L'objet de la science politique serait ainsi d'étudier cette politisation³,

¹ Bruno Latour définit la *politics* comme « *the progressive composition of a common world* ».

Bruno Latour, *Reassembling the Social. An introduction to Actor-Network Theory*, Oxford, Oxford University Press, 2005, p. 254.

² Pierre Favre, « Ce que les science studies font à la science politique. Réponse à Bruno Latour. », *Revue française de science politique*, 2008, vol. 58, n° 5, p. 819.

³ Sur la notion de politisation en science politique, voir l'ouvrage de Jacques Lagroye, *La politisation*, Paris, Belin, 2003, 576 p.

soit le « *processus social de saisie du champ politique*¹ ». La délimitation de ce champ politique n'est pas une évidence comme le soulignent les récurrents débats à ce sujet².

La perspective non essentialiste défendue par Latour, qui refuse toute définition a priori d'un domaine qui relèverait du « politique » pour s'intéresser aux manifestations, aux « états³ » des causes, aux différentes formes de politique qu'elles peuvent prendre, prend alors tout son sens⁴. Dans cette optique, les *politiques de la donnée* correspondent alors aux différents états des données dans leur modalité de composition d'un monde commun. Les politiques sont considérées comme des étapes distinctes dans la trajectoire des mêmes données. Ces politiques de la donnée peuvent s'inscrire dans ce qui relève d'un champ politique pour la science politique, mais elles ne s'y restreignent pas. Les données peuvent prendre la forme d'une politique publique, être débattues au sein de conseils municipaux, être l'objet de mobilisation sociale, etc. Cependant, elles sont également produites par des habitants dans leurs pratiques quotidiennes, circulent dans des réseaux de télécommunications, sont stockées dans des *data centers* à l'autre bout du monde ou utilisées par des entreprises dans leurs processus de production. Difficile alors de réduire les données à une politique. Il existe plutôt autant de politiques qu'il y a d'étapes dans les transformations des données. La composition du réseau qui se déploie autour de la donnée varie selon les traductions dont elle est l'objet. Le spectre de la question ne se réduit pas à la *politique publique des données*, au singulier, mais bien aux *politiques des données*, au pluriel.

¹ P. Favre, « Ce que les science studies font à la science politique. Réponse à Bruno Latour. », art cit, p. 820.

² Voir par exemple le numéro spécial de la revue *Espaces Temps* coordonné par Sophie Duchesnes et Florence Haegel, ou les contributions de Pierre Favre et Bernard Voutat dans l'ouvrage *L'atelier du politiste*. Sophie Duchesne et Florence Haegel (eds.), « Repérages du politique. Regards disciplinaires et approches de terrain », *Espaces Temps*, 2001, vol. 76, n° 1, p. 1-154 ; Pierre Favre, « La question de l'objet de la science politique a-t-elle un sens ? » dans Pierre Favre, Olivier Fillieule et Fabien Jobard (eds.), *L'atelier du politiste. Théories, actions, représentations.*, Paris, La Découverte, 2007, p. 17-33 ; Bernard Voutat, « A propos de l'objet de la science politique : sens et non-sens d'une question récurrente » dans Pierre Favre, Olivier Fillieule et Fabien Jobard (eds.), *L'atelier du politiste. Théories, actions, représentations.*, Paris, La Découverte, 2007, p. 35-57.

³ « « Science » et « politique » ne qualifient par des domaines d'activité ou de compétence, mais plutôt des états – comme on parle d'état liquide ou d'état gazeux - dans lesquels se trouvent les objets controversés qui ont vocation à composer le monde commun. »

Bruno Latour, « Pour un dialogue entre science politique et science studies », *Revue française de science politique*, 2008, vol. 58, n° 4, p. 677.

⁴ La conception du politique de Bruno Latour a évolué au long de sa carrière : Graham Harman en distingue trois, Ève Seguin deux. Nous retiendrons ici la première conception que Latour a exprimé du politique comme modes de composition d'un monde commun. Graham Harman, *Bruno Latour: Reassembling the Political*, London, Pluto Press, 2014, 216 p ; Ève Seguin, « Pourquoi les exoplanètes sont-elles politiques ? Pragmatisme et politicalité des sciences dans l'œuvre de Bruno Latour », *Revue française de science politique*, 2015, vol. 65, n° 2, p. 279-302.

Tout l'enjeu est alors de parvenir à suivre les épreuves et les différents états de transformation par lesquels passent les données. Latour – à l'instar de nombreux travaux de sociologie des sciences et des techniques – préconise pour cela de s'intéresser aux moments controversés, aux affaires, aux *issues*, au cours desquels les êtres sont requalifiés. En mettant en circulation les données, l'*open data* constitue un moment privilégié pour étudier les politiques des données.

IV. L'open data : la mise en circulation des données urbaines

L'*open data* est un exemple paradigmatique de l'idéologie du libéralisme informationnel appliqué au gouvernement urbain. Ce mouvement, initié dans les pays anglo-saxons, vise à mettre en circulation les données des administrations publiques afin de les rendre accessibles aux citoyens. En France, la ville de Rennes est la première collectivité à s'engager dans une politique d'*open data*. La mise à disposition de ses données doit favoriser à la fois le développement économique et la transparence et la participation citoyenne. Tant des entrepreneurs que des citoyens sont invités à s'emparer de ces données. Rennes sera suivi par Paris (27 janvier 2011), Montpellier (mars 2011) Toulouse (22 octobre 2011) ou Nantes (21 novembre 2011). Aujourd'hui, plus d'une centaine de collectivités sont engagées dans des démarches d'ouverture de leurs données¹. Ce chiffre devrait s'étendre dans les prochains mois puisque la loi NOTRE², votée en 2015, impose l'ouverture des données à l'ensemble des collectivités de plus de 3500 habitants.

A plusieurs titres, l'*open data* est un cas d'étude particulièrement intéressant pour analyser les politiques de données urbaines. A l'instar des situations de controverses³ ou des

¹ 106 collectivités qui se répartissent ainsi : 10 régions, 19 départements, 29 EPCI et 48 municipalités.

Source : <http://www.opendatafrance.net/lassociation/les-acteurs-2> (Consulté le 28 janvier 2016)

² Loi n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République

³ Michel Callon, « Pour une sociologie des controverses technologiques » dans Madeleine Akrich, Michel Callon et Bruno Latour (eds.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'École des Mines, 2006, ; Cyril Lemieux, « À quoi sert l'analyse des controverses ? », *Mil neuf cent*, janvier 2007, n° 25, p. 191-212.

pannes techniques¹, l'ouverture des données constitue des moments d'épreuve qui présentent l'avantage d'obliger à l'explicitation de ce qui reste enfoui dans les routines ordinaires. L'*open data* impose une ouverture de la boîte noire des données. Le travail préalable à la diffusion des données oblige les collectivités à un travail de réflexivité sur les données. De quelles données dispose-t-on ? Qui les utilisent ? Quelles sont les représentations qu'elles véhiculent ? Que peut-on en faire ?

En outre, la mise en circulation des données qui est au cœur de l'*open data* conduit à recomposer leur réseau sociotechnique. Pour passer d'une logique de stock à une logique de flux, les données doivent être détachées de leur environnement initial pour être attachées à de nouveaux acteurs (citoyens, organisations, entreprises, etc.). Pour cela, l'identité des données, ainsi que celle de chacun des acteurs doit être traduite, pour parvenir à construire et stabiliser ces nouveaux réseaux sociotechniques.

Enfin, l'*open data* se caractérise par son indétermination. L'incertitude règne lors de la construction de ces nouveaux réseaux. De quelles données parle-t-on ? Qui s'en emparera ? Comment les mettre à disposition ? Personne n'a de réponses précises à apporter à ces questions. Les acteurs tâonnent, explorent, émettent des inquiétudes et des réticences, refusent de s'inscrire dans ces réseaux recomposés, etc. Suivre la politique d'*open data* d'une collectivité révèle toutes ces situations quotidiennes d'incertitude. Une indétermination fondamentale règne, ce qui permet de mesurer l'étendue du champ des possibles qui s'offre aux acteurs. L'enjeu est alors d'étudier finement les cadrages et recadrages qui ont lieu dans ces situations pour comprendre comment les identités des actants sont stabilisées ou au contraire remises en cause.

L'*open data* est ainsi un cas d'étude privilégié pour analyser les politiques des données qui émergent de la mise en circulation de ces dernières. L'*open data* est une épreuve : un moment d'incertitude où rien n'est figé. Alors que le champ des possibles est ouvert, on s'interroge pêle-mêle sur les données, leur producteur, leurs utilisateurs, les conséquences en terme de gouvernement, les infrastructures techniques à mettre en œuvre, etc. Alors que les choix ne sont pas encore effectués, cette thèse met en évidence la pluralité des agencements

¹ Susan Leigh Star et Karen Ruhleder, « Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces », *Information Systems Research*, 1996, vol. 7, n° 1, p. 111-134.

qui émergent de ces solutions envisagées, qui constituent autant de politiques des données urbaines.

V. Méthodologie

La sociologie des données poursuivie dans cette thèse est appréhendée par la mise en œuvre d'une démarche ethnographique qui offre un accès privilégié aux espaces de production et d'utilisation de ces données. Le recours à l'enquête ethnographique satisfait à un parti pris empirique visant à étudier, en situation, les acteurs, humains ou non humains. Effectuer une ethnographie des données peut sembler étrange si on les réduit à des entités purement techniques. Pourtant, leur boîte noire entrouverte, les données deviennent des systèmes sociotechniques, dotées d'une « propriété relationnelle »¹. Tout l'enjeu d'une sociologie des données est, comme le formule Susan Leigh Star, de parvenir à « *faire remonter à la surface le travail invisible*² » nécessaire à la conception et au maintien de ces systèmes. Le recours à une démarche ethnographique, qui se caractérise par une insertion personnelle de longue durée au sein du milieu étudié³, autorise l'accès à des arènes où les choix techniques sont analysés, déconstruits, arbitrés⁴. Pour cette thèse, cette observation ethnographique a été effectuée au sein de la Métropole de Lyon dans le cadre d'un dispositif CIFRE. Les conventions industrielles de formation par la recherche sont un dispositif de financement de thèse, qui lie un organisme (public ou privé), un laboratoire de recherche et un doctorant. Ce dernier partage son temps de travail entre l'entreprise – dans mon cas une administration territoriale – et le laboratoire. En contrepartie, l'organisme, qui reçoit une subvention annuelle de l'Association nationale de la recherche et de la technologie (ANRT), verse un salaire au doctorant. Le chercheur est ainsi non seulement engagé *dans* mais

¹ *Ibid.*, p. 113.

² Susan Leigh Star, « The Ethnography of Infrastructure », *American Behavioral Scientist*, 1999, vol. 43, n° 3, p. 385.

³ Olivier Schwartz, « L'empirisme irréductible » dans *Le hobo : sociologie du sans-abri*, Paris, Nathan, 1993, p. 265-308.

⁴ Les données sont un objet technique, étudié dans cette thèse par un chercheur en sciences sociales. Ce regard externe offre de nouveaux points de vue, toutefois, des erreurs ou des inexactitudes techniques ont pu se glisser dans l'analyse.

également *par* son terrain¹. Ce financement de la recherche doctorale n'est pas sans conséquence sur le terrain étudié, les méthodes employées et l'objet analysé.

A. Le doctorant CIFRE : défis et opportunités du chercheur acteur de son objet

“*The most difficult thing about doing fieldwork is remembering who you are*”².

Cette citation attribuée à Erving Goffman, s'applique parfaitement au doctorant CIFRE, dont la double appartenance de professionnel et de chercheur suscite de nombreux questionnements identitaires, éthiques et épistémologiques. Plutôt que de passer sous silence ces doutes, je souhaite revenir dans cette introduction sur le positionnement spécifique de la CIFRE qui participe indubitablement à la construction et à l'analyse de l'objet de recherche³. S'inscrivant dans un idéal scientifique de réflexivité⁴, ce questionnement épistémologique, mené tout au long de la thèse, vise à rechercher la « bonne » distance quant à l'engagement, tout en acceptant d'être affecté par son terrain.

1) Du « Eux » au « Nous ». Entre observation et participation : travailler sur l'institution et pour l'institution

La situation CIFRE offre un énorme avantage pour la recherche puisqu'elle permet de s'insérer dans un univers professionnel, souvent difficile d'accès, afin de l'étudier. En offrant au doctorant CIFRE un statut de professionnel, intégré dans une équipe de travail, elle ouvre la porte de milieux professionnels, et surtout, elle tend à minorer le statut de chercheur du doctorant vis-à-vis des acteurs, qui deviennent ses collègues. Le chercheur est rapidement davantage considéré comme un collègue avec lequel on travaille plutôt qu'un observateur extérieur. Néanmoins, du chercheur devenu collègue est attendue une participation classique

¹ Sur des questionnements similaires, voir notamment : Thomas Alam, Marion Gurruchaga et Julien O'Miel, « Science de la science de l'État : la perturbation du chercheur embarqué comme impensé épistémologique. », *Sociétés contemporaines*, 2012, vol. 3, n° 87, p. 155-173.

² Citation attribuée à Erving Goffman par sa fille Alice Goffman. Source : <http://www.nytimes.com/2016/01/17/magazine/the-trials-of-alice-goffman.html> (Consulté le 29 mars 2016)

³ T. Alam, M. Gurruchaga et J. O'Miel, « Science de la science de l'État : la perturbation du chercheur embarqué comme impensé épistémologique. », art cit ; Julie Landour, « Le chercheur funambule. Quand une salariée se fait la sociologue de son univers professionnel », *Genèses*, 2013, vol. 1, n° 90, p. 25-41 ; Olivia Foli et Marlène Dulaurans, « Tenir le cap épistémologique en thèse Cifre. Ajustements nécessaires et connaissances produites en contexte », *Etudes de communication*, 2013, n° 40, p. 59-76 ; Gérald Gaglio, « En quoi une thèse CIFRE en sociologie forme au métier de sociologue ? Une hypothèse pour ouvrir le débat », *Socio-logos*, 2008, n° 3.

⁴ Pierre Bourdieu, *Le sens pratique*, Paris, Les Editions de Minuit, 1980.

au travail. Le doctorant CIFRE est ainsi soumis à la tension entre observation et participation, devenue réelle implication dans l'objet étudié. En effet, en participant activement à l'ordre de travail, il est impossible de garder une posture d'observateur qui demande idéalement une distanciation quant à la situation. Comment alors récolter des données situationnelles pour le travail de recherche ? Comment à la fois « en être » et pouvoir « s'en défaire »¹ ? La posture du chercheur CIFRE met en évidence le double mouvement de l'attachement et du détachement qui caractérise la science, contrainte de s'attacher à certains acteurs puis de s'en détacher pour produire des connaissances². Elle s'observe dans les modifications successives de mon implication sur le terrain de recherche : d'une posture de chercheur observateur dans les premiers mois de la CIFRE, un processus « d'indigénisation » m'a conduit à devenir un professionnel engagé dans l'action, devant, par la suite, se distancier de ce collectif pour produire des connaissances scientifiques.

Avant même de débiter la CIFRE, les premiers contacts avec les agents de la communauté urbaine m'ont fait sentir le travail nécessaire pour parvenir à appartenir simultanément aux mondes universitaire et professionnel. En effet, les questionnements et le vocabulaire mis en avant dans les premiers projets de thèse étaient loin de faire écho aux préoccupations des agents territoriaux. Dès lors, pour susciter leur intérêt à financer cette thèse, il m'a été nécessaire d'effectuer un travail sémantique de choix de termes faisant sens tant pour la recherche académique qu'au sein des collectivités³. Ce travail d'insertion dans le milieu professionnel d'une administration territoriale s'est poursuivi lors de mon entrée « effective » sur le terrain étudié. Les premiers mois, qui ont suivi mon recrutement comme chargé de mission « open data » au sein de la communauté urbaine de Lyon, ont été une intense période de socialisation à l'institution, à ses codes, son langage et son fonctionnement. Cette phase de découverte a été riche en étonnements fertiles pour la thèse. Elle a également été une phase de mise à l'épreuve de mon statut d'étudiant, nécessaire pour me faire une place et m'attacher à ce milieu professionnel. Pour maîtriser les interactions quotidiennes, il m'a

¹ Elodie Lemaire, « “En être” et “s'en défaire”. Retour réflexif sur une enquête par immersion dans la police », Rennes, 2013.

² Michel Callon, « Ni intellectuel engagé, ni intellectuel dégage : la double stratégie de l'attachement et du détachement », *Sociologie du travail*, 1999, n° 41, p. 65-78.

³ C'est le cas par exemple des termes « gouvernance » ou « instruments », qui font sens pour les agents administratifs, même s'ils ne leur assignent pas la même signification que dans le milieu universitaire. Il est toutefois important de préciser que malgré ces « traductions » initiales (de mon propre fait), les personnes chargées de mon recrutement, puis de mon suivi, n'ont jamais questionné le fond de mon projet de thèse. J'ai bénéficié d'une totale liberté quant à la construction et la redéfinition de celui-ci.

fallu notamment acquérir la langue « indigène ». Ce jargon, composé de sigles et de références qui ne font pas sens, constitue autant de normes de langage permettant d'ériger une barrière entre ceux qui sont dans l'institution et ceux qui en sont extérieurs¹. La découverte du milieu professionnel et la découverte du terrain d'étude ne font qu'un. Aussi, elle s'est accompagnée d'une phase de récolte de données par la découverte de documents de travail internes, des entretiens informels exploratoires et l'observation scrupuleusement reportée dans un journal de terrain.

Cette insertion au sein de la communauté urbaine a été facilitée par le développement de compétences utiles à l'organisation. J'ai réussi à me faire une place au sein de l'institution en mobilisant des savoir-faire académiques au service du projet d'ouverture des données. L'expertise sur l'*open data* acquise dans le milieu universitaire s'est traduite par des notes de synthèse et la production de recommandations. Ayant de vagues connaissances en droit, j'ai également été associé à un travail juridique conséquent alors que ces compétences manquaient au sein de l'équipe projet. En mobilisant différentes aptitudes, j'ai réussi à jouer avec mes différentes identités, me positionnant tour à tour en tant que chercheur, chargé de mission ou expert, pour démultiplier les situations d'observations. Virant peu à peu « indigène² », cet engagement, loin de devenir « problématique³ », a enrichi considérablement ma recherche et le projet d'*open data* de la communauté urbaine de Lyon. Pour reprendre l'expression de Donna Haraway⁴, je suis *devenu avec* mon terrain d'étude. Je me suis laissé affecté par mon travail quotidien au sein du Grand Lyon tout comme ce travail a été affecté par mes travaux de recherche. La restitution régulière de mes réflexions académiques, sous différentes formes (réunions, rapports, présentations, discussions, etc.), à mes collègues du Grand Lyon, a influé sur le projet *open data* de la collectivité. Plus généralement, par ma posture de chercheur impliqué dans le projet, j'ai participé à la reconfiguration du projet, en faisant émerger d'autres entités, en rendant visible des liens non visibles, par mes questions, mes remarques et mon implication.

¹ Tout comme a contrario il m'a fallu prendre conscience et « oublier » le jargon universitaire, incompréhensible pour les praticiens.

² Symboliquement, le « nous » a remplacé peu à peu le « ils » dans mes notes de terrain

³ Valérie Amiraux et Daniel Cefaï, « Les risques du métier. Engagements problématiques en sciences sociales. », *Cultures & Conflits*, 2002, n° 47.

⁴ Donna Haraway, *When Species Meet*, Minneapolis, University of Minnesota Press, 2008, 423 p.

Les frontières entre la recherche et l'action sont ainsi indissociablement brouillées. Comment alors réussir à produire des connaissances scientifiques si le projet observé et le travail de recherche sont influencés l'un par l'autre ? Comme l'a affirmé Michel Callon¹, il est nécessaire de se détacher des acteurs de ce collectif pour parvenir à un degré de généralité qui rende la connaissance suffisamment robuste pour être transportée au sein d'autres univers sociaux. Pour parvenir à me distancier de la singularité du cas étudié, j'ai procédé à un double déplacement géographique et temporel. J'ai effectué un séjour de recherche à l'université de Columbia, durant lequel j'ai pu couper avec le quotidien de la communauté urbaine de Lyon, tout en effectuant un travail de comparaison. Au retour de ce séjour, j'ai réalisé un travail d'archive afin d'inscrire l'observation ethnographique dans un temps plus long. J'ai également mené une série d'entretiens approfondis. Enfin, la lecture a posteriori de mon journal de terrain a été très bénéfique pour mesurer les évolutions du cas observé et mener l'analyse préalable à la rédaction. En résumé, ma « politique du terrain² » s'est basée sur des sources multiples pour assurer une distanciation : l'observation participante, soigneusement recensée dans un journal de terrain, des sources écrites (documents internes et archives), des entretiens compréhensifs et un décentrement comparatif.

2) Rendre compte de l'enquête Cifre : le choix du récit ethnographique

En ethnographie, « l'expérience incarnée de l'enquêteur » est le principal médium de l'enquête comme le souligne Daniel Céfai³. Dès lors, plutôt que d'essayer de cacher mes propres conceptions et mon implication sur le terrain étudié, j'ai décidé de les mettre en avant et de les utiliser comme une ressource centrale de l'analyse. Plutôt que de nier ma double appartenance académique et professionnelle, cette thèse vise à rendre compte de la recherche menée en tant que doctorant – chargé de mission au sein de la Métropole de Lyon. Cette implication personnelle se traduit dans le manuscrit de thèse par l'emploi de la première personne du singulier ainsi que par l'utilisation du journal de terrain comme un matériau d'enquête.

¹ M. Callon, « Ni intellectuel engagé, ni intellectuel dégage : la double stratégie de l'attachement et du détachement », art cit.

² Jean-Pierre Olivier de Sardan, « La politique du terrain. Sur la production des données en anthropologie. », *Enquête*, 1995, n° 1, p. 71-109.

³ Daniel Céfai, « Introduction. L'engagement ethnographique. » dans Daniel Céfai (ed.), *L'engagement ethnographique*, Paris, Editions de l'EHESS, 2010, p. 7.

La thèse est un récit rendant compte d'une perspective particulière sur les données et leur ouverture. Le propre des approches ethnographiques est de ne pas viser à l'exhaustivité, mais de proposer une mise en récit de l'objet observé. L'enquête qualitative peut difficilement prétendre à la représentativité. Cette mise en récit est liée au terrain choisi, mais également à la place sur ce terrain. J'ai un point d'observation au sein de cette politique d'ouverture de données qui est celui du chargé de mission au sein de l'équipe projet. Cela m'a permis d'avoir accès à certains points de vue, mais également d'être hermétique à d'autres. Si j'ai essayé de limiter les biais de cet ancrage dans le terrain par des entretiens ainsi que par une comparaison limitée du cas lyonnais avec d'autres villes nord-américaines, la subjectivité de l'analyse est assumée. Comme le souligne avec humour Clifford Geertz : « *Au pays des aveugles, qui ne sont pas si peu observateurs qu'ils en ont l'air, le borgne n'est pas roi, il est spectateur*¹ ». L'ethnographie est toujours une interprétation. La plausibilité de celle-ci réside dans le reflet de l'enquête empirique que donne à voir l'ethnographe à son lecteur. Dans cette thèse, ce « pacte ethnographique² » se traduit par la mobilisation *in extenso* d'extraits d'entretiens, de notes de terrain et de documents internes analysés. Dans ce travail d'explicitation se trouve la possibilité de la généralisation : à partir de la description dense (*thick description*) d'un cas singulier, le compte-rendu « transporte³ » la situation locale et rend possible la fabrication d'équivalence à d'autres lieux et d'autres acteurs.

B. L'étude d'un cas : la Métropole du Grand Lyon

L'ethnographie est située dans le temps et dans l'espace. La délimitation spatiale et temporelle de mon terrain d'enquête et d'analyse a été déterminée par le dispositif de financement de cette thèse. La convention CIFRE a débuté en novembre 2012. J'ai alors rejoint la communauté urbaine de Lyon en tant que chargé de mission *open data*. J'ai suivi la mise en place et l'évolution de cette politique de mise à disposition de données jusqu'en juin 2015. En tant que chargé de mission, j'ai pu suivre cette politique *en train de se faire* en accédant aux coulisses de l'administration et aux arènes discrètes difficilement perceptibles par des sources externes (entretiens, archives) ou par une ethnographie plus courte.

¹ Clifford Geertz, « Du point de vue de l'indigène : sur la nature de la compréhension anthropologique. » dans *Savoir local, savoir global. Les lieux du savoir.*, Paris, PUF, 1986, p. 75.

² J.-P. Olivier de Sardan, « La politique du terrain. Sur la production des données en anthropologie. », art cit.

³ M. Callon, « Ni intellectuel engagé, ni intellectuel délogé : la double stratégie de l'attachement et du détachement », art cit.

La spécificité de cet accès au terrain a empêché sa reproduction au sein d'autres institutions. La thèse se centre ainsi essentiellement sur le cas de la communauté urbaine de Lyon. La communauté urbaine de Lyon est une structure intercommunale, créée en 1969 et regroupant 59 communes. Elle dispose de compétences de services urbains (eau, assainissement, transport, propreté), d'aménagement urbain et de développement économique. Le 1^{er} janvier 2015, le Grand Lyon a fusionné, sur son territoire, avec le conseil général du Rhône, pour former la Métropole de Lyon¹. Aux compétences initiales, se sont ajoutées celles issues du département du Rhône (insertion, personnes âgées et handicapées, enfance, éducation, culture, etc.)².

La thèse se concentre sur l'étude privilégiée d'un cas. La question de la représentativité de cet objet empirique est un leurre dans la mesure où, pour y répondre, il faudrait réduire la complexité inhérente de ce cas à des critères permettant de le rapporter à une totalisation. L'enjeu de la pensée par cas réside davantage dans sa généralisation. L'étude de cas procède « *par l'exploration et l'approfondissement d'une singularité accessible à l'observation [pour] en extraire une argumentation de portée plus générale, dont les conclusions seront réutilisables* »³. Penser par cas c'est ainsi raisonner à partir de singularités pour faire émerger des questions de portée théorique générale. Si le terrain étudié est irréductible, il n'en est pas de même des processus et des phénomènes qu'il donne à voir.

En atteste le travail de comparaison menée dans quelques villes nord-américaines (New York, Boston, Philadelphie) qui m'a permis de confirmer les premiers résultats émergents de mon terrain lyonnais. L'étude des politiques d'*open data* de ces villes repose sur l'analyse de documents produits par les administrations, ainsi que sur une dizaine d'entretiens compréhensifs auprès d'acteurs en charge de ces politiques. Cette comparaison asymétrique m'a permis de confirmer et nuancer mes premières hypothèses de recherche. Toutefois, l'asymétrie de l'ampleur du travail de récolte de données m'a conduit à baser le récit sur le

¹ Sur les dynamiques de construction du rôle central de la communauté urbaine de Lyon dans le gouvernement métropolitain et les défis auxquels va être confrontée la nouvelle institution, voir Deborah Galimberti et al., « La métropole de Lyon. Splendeurs et fragilités d'une machine intercommunale. », *Hérodote*, 2014, vol. 3, n° 154, p. 191-209.

² Le changement institutionnel a été effectif durant le terrain d'étude. Il n'a toutefois pas eu d'effets directs sur la politique d'ouverture des données, si ce n'est une réorganisation administrative mentionnée dans le chapitre 5. C'est pourquoi dans cette thèse, par commodité, il est fait référence tout à la fois à la communauté urbaine de Lyon, au Grand Lyon ou à la Métropole de Lyon pour désigner l'institution administrative, sans que la signification n'ait d'importance pour l'analyse.

³ Jean-Claude Passeron et Jacques Revel, *Penser par cas*, Paris, Editions de l'EHESS, 2005, p. 9.

cas lyonnais, tout en le mettant en perspective par des exemples nord-américains sur des points précis de l'argumentation.

C. Une étude centrée sur des données spatialisées

Les données étudiées dans cette thèse sont des données spatialisées. Ce choix se justifie par leur rôle crucial dans le développement des *smart cities*. La carte est devenue l'interface privilégiée de correspondance entre les mondes physiques et numériques. Elle constitue, selon Antoine Picon, l'essence même des villes intelligentes. « *Ce n'est pas tant la forme de la ville qui change que sa carte, ou plutôt ses cartes, tant l'avènement du numérique coïncide avec une prolifération de représentations cartographiques du phénomène urbain. C'est au travers de cette prolifération que la ville devient intelligente* »¹. Ce renouveau de la cartographie, devenue dynamique et interactive, repose sur plusieurs évolutions technologiques qui ont profondément transformé les représentations et les pratiques urbaines². Le développement des services de cartographie en ligne et l'automatisation de la géolocalisation ont contribué à l'utilisation de l'information géographique auprès d'un large public. Toutefois, ces nouveaux usages de la cartographie sont permis grâce à l'essor des systèmes d'information géographique (SIG) dans l'ensemble des secteurs urbains³. La contribution des SIG au foisonnement de l'information spatialisée est essentielle puisque ces outils, reposant sur des bases de données relationnelles, associent à chaque entité représentée une référence spatiale localisante. Les cartes ne sont qu'une modalité d'utilisation de ces données. Ce sont ces dernières qui sont au cœur du renouvellement des représentations et des pratiques cartographiques.

¹ A. Picon, *Smart Cities. Théorie et critique d'un idéal auto-réalisateur.*, op. cit., p. 101.

² Boris Beaudé, « Espace de la ville, espace de la carte; de l'analogie à la coexistence » dans Khaldoun Zreik (ed.), *Nouvelles cartographies, nouvelles villes HyperUrbain.2*, Paris, Europa productions, 2010, p. 15-39.

³ Au point que l'on peut s'interroger : existe-t-il des données urbaines non géographiques ? Si la réponse est positive, de nombreuses données font l'objet d'un processus de spatialisation. A titre d'exemple, les données « sociales » de la Métropole de Lyon ne sont aujourd'hui pas intégrées au sein d'un système d'information géographique. Toutefois, cette évolution est envisagée afin de faciliter leur croisement avec d'autres données.

VI. Suivre la donnée : Les étapes du récit

A partir du cas de l'*open data*, cette thèse a pour ambition d'analyser au concret la mise en circulation des données urbaines. Pour dépasser les discours sur la *smart city* qui considèrent les données comme neutres, autonomes et aisément circulables, j'effectue une sociologie des données attentive aux transformations conjointes des données et des acteurs qui leur sont associés. Cette approche met en évidence une tension entre attachement et détachement lorsque l'on souhaite passer d'une logique de stock à une logique de flux de données. Cette tension constitue le fil directeur de cette thèse, organisée en cinq chapitres, qui sont autant d'étapes de la chaîne des données ouvertes : produire, mettre en circulation, ouvrir, réutiliser et gouverner¹. A chacune de ces étapes, la thèse pointe les politiques des données urbaines, entendues comme les modes d'association de collectifs autour des données.

Le premier chapitre s'intéresse aux processus de production des données géographiques et mobilité de l'agglomération lyonnaise. Il met en évidence le fait que la donnée est constituée de multiples attachements qui la rendent indissociable des vastes réseaux sociotechniques dans lesquels elle est insérée. Attachée, la donnée est également attachement. Ces liens la dotent d'un pouvoir d'agir, d'une *agency* : en mettant en relation d'une façon particulière les utilisateurs et l'espace urbain, elle participe à l'émergence d'agencements associant l'institution communautaire à son territoire.

Le deuxième chapitre porte sur la problématisation de la politique d'*open data*. Il analyse comment la mise en circulation des données s'est imposée au sein de la communauté urbaine de Lyon et comment ce libéralisme informationnel a été traduit pour répondre aux politiques communautaires. Ce chapitre retrace le cadrage spécifique de la politique de diffusion de données du Grand Lyon qui met en évidence des volontés d'attachement particulier de ces données à de nouveaux utilisateurs.

Le troisième chapitre étudie les opérations nécessaires à la mise en circulation des données. La donnée subit une série d'épreuves de diffusibilité qui participent simultanément à son détachement de son environnement initial et à son attachement à de nouveaux utilisateurs.

¹ Il convient de préciser que ces cinq étapes ne forment pas un parcours linéaire de la donnée de sa production à son gouvernement. Les va-et-vient entre les différentes étapes sont constants.

La donnée est modifiée tout au long de la chaîne de diffusion pour acquérir des propriétés de *diffusibilité* et se stabiliser comme une entité diffusable. Ce processus est réversible : des données non catégorisées comme telles peuvent le devenir, et, inversement, des données ouvertes peuvent être redéfinies comme non-diffusables.

Le quatrième chapitre interroge les modalités de réutilisation des données mises à disposition. Il met en avant le travail nécessaire à l'attachement de ces données à de nouveaux utilisateurs. Pour dépasser leur cadrage originel, il est nécessaire de traduire ces données sans nuire à leur solidité. Au travers de l'étude de plusieurs réutilisations de données mises à disposition par le Grand Lyon, le quatrième chapitre révèle l'existence de différentes politiques de réutilisation. Présentées dans trois catégories selon les opérations qu'elles font subir aux données (la consolidation, l'homogénéisation et l'articulation), ces modalités plurielles d'attachement des données à de nouveaux utilisateurs font émerger de nouveaux agencements influant sur le gouvernement de la ville.

Le cinquième chapitre analyse les modes de gouvernement des flux de données issus de leur mise en circulation. A partir du programme d'*open data*, la direction des services informatiques du Grand Lyon tente d'ériger la donnée en politique publique autonome dont elle assurerait le pilotage à l'échelle de la Métropole. Pour cela, elle souhaite se positionner en situation d'intermédiaire entre producteurs et utilisateurs de données. Toutefois, les difficultés rencontrées par cette direction pour réguler la circulation des données révèlent une tension latente entre autonomisation et sectorialisation de cet enjeu de la donnée.

Chapitre 1

Produire : Attachements des données et agencements territoriaux

Ce premier chapitre traite de la mise en données de la ville. La *mise en données* (« *datafication*¹ ») est un processus de quantification, qui vise à « *exprimer et faire exister sous une forme numérique ce qui, auparavant, était exprimé par des mots et non par des nombres* »². Cette quantification s'est étendue bien au delà des enquêtes statistiques et des registres étudiés par Alain Desrosières³. Le phénomène de numérisation croissante du monde, soit la conversion d'informations analogiques en code binaire manipulable par des ordinateurs, amplifie la quantité de données aujourd'hui produites. La donnée est en effet au cœur des systèmes informatiques, qui nous aident à gérer de l'information représentée sous forme numérique⁴. Si l'on s'en tient à une conception informatique, la donnée est un élément binaire, représenté par une séquence de *bits*⁵, qui peut circuler, être stocké et être analysé. Outre sa numérisation, le caractère essentiel de la donnée informatique serait son immatérialité : elle n'est pas liée à un support physique ce qui facilite, a priori, sa duplication et sa circulation⁶.

Encadré 1 : Données, jeu de données, base de données

Les données sont souvent rassemblées dans des jeux de données (*dataset*). Un *dataset* est un ensemble de données. Il correspond généralement à un tableau dans lequel les colonnes représentent une variable particulière, un attribut, et les lignes une entité observée. Chaque intersection de ce tableau est une donnée.

Afin de faciliter leur gestion et leur interrogation, ces jeux de données sont eux-mêmes insérés dans des bases de données. Celles-ci sont des outils, aujourd'hui informatisés, qui

¹ La « datafication » est un néologisme forgé par Mayer-Schönberger et Cukier. Dans la version française de leur ouvrage, il est traduit par « mise en données » qu'ils définissent ainsi : « *Mettre en données un phénomène consiste à le transposer en un certain nombre de chiffres que l'on va pouvoir mettre en tableau, puis analyser.* » (p. 96)

Viktor Mayer-Schönberger et Kenneth Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, New York, Houghton Mifflin Harcourt, 2013, 240 p.

² Alain Desrosières, *Pour une sociologie historique de la quantification*, Paris, Presses de l'Ecole des mines, 2008, p. 10.

³ Alain Desrosières, « Décrire l'Etat ou explorer la société : les deux sources de la statistique publique », *Genèses*, 2005, n° 58, p. 4-27.

⁴ Gérard Berry, *Pourquoi et comment le monde devient numérique*, Paris, Fayard, 2008, 78 p.

⁵ Un *bit* est une variable binaire pouvant prendre la valeur 0 ou 1.

⁶ Ce caractère immatériel de la donnée est questionné dans la suite de ce chapitre où, au contraire, la matérialité de la donnée est mise en évidence. La donnée s'incarne dans tout un ensemble de personnes, de capteurs, de réseaux, d'ordinateurs, de *data centers*, etc., qui sont on ne peut plus matériel, et sans lesquels la donnée n'aurait pas d'existence.

permettent de rassembler l'information selon une organisation structurée. Les données peuvent être organisées de multiples manières et leur structuration dans ces bases est cruciale pour déterminer l'usage qui pourra en être fait¹. « Une base de données est donc une collection de données reliées entre elles par des relations. Elle permet l'interrogation et la manipulation de ces données par des langages de haut niveau². » Elles sont mises en forme par un système de gestion de base de données (SGBD) qui permet également leur manipulation et leur information. Le SGBD est un dispositif assurant la médiation entre l'individu et l'ensemble des données. Il propose des « vues » différentes qui offrent plus ou moins de possibilité d'interaction avec la base de données. Un modèle très courant de base de données est le modèle relationnel. Mis au point dans les années 70 par Ted Codd, le modèle relationnel organise les données en tableaux à deux dimensions appelées *tables* (ou *relations*). Chaque ligne de ces tables, une entité et ses attributs, est dénommée *n-uplet*, et l'ensemble des *n-uplets* d'une table est une *relation*. La popularité de ce modèle de base de données réside dans la puissance du langage, reposant sur l'algèbre relationnel, qui permet des requêtes simples, logiques et précises³.

La majorité des données présentes dans une institution ne sont néanmoins pas mises en base, elles sont stockées directement par l'application dans les fichiers qu'elle manipule. Une base de données se distingue ainsi d'un fichier informatique, qui contient des informations numériques (des données) réunies sous un même nom et enregistrées selon un format⁴ particulier sur un support de stockage. Ces fichiers contiennent des données dites non-structurées, c'est-à-dire qui n'ont pas de structure identifiée ou de modèle défini⁵.

Dans son histoire étymologique du terme « data », Daniel Rosenberg souligne que la fonction initiale de la donnée est rhétorique. « *Facts are ontological, evidence is*

¹ Nous reviendrons plus longuement sur cette question essentielle dans la suite de cette thèse. BOWKER G.C., *Memory Practices in the Sciences*, MIT Press, 2005 ; RUPPERT E., « The Governmental Topologies of Database Devices », *Theory, Culture & Society*, 2012, vol. 29, n° 4-5, p. 116-136.

² Gilles Dowek, *Introduction à la science informatique*, Paris, Centre régional de documentation pédagogique de l'académie de Paris, 2011, p. 312.

³ Serge Abiteboul, *Sciences des données: De la logique du premier ordre à la Toile*, Paris, Fayard, 2012, 112 p.

⁴ Une convention définissant comment les informations et les métadonnées sont numérisées et séquencées dans le fichier

⁵ Par commodité et pour être fidèle au discours des acteurs, sauf mention contraire, j'utilise dans cette thèse le terme générique de « données » pour qualifier indistinctement les « données », « jeux de données » ou « bases de données ».

*epistemological, data is rhetorical*¹. La donnée a un statut pré-factuel : elle est un élément fondamental, considéré comme acquis, sur lequel se base un raisonnement ou une argumentation. Contrairement au fait, la donnée est *donnée*, sa véracité n'a pas à être démontrée. *“When a fact is proven false, it ceases to be a fact. False data is data nonetheless”*². L'usage historique et la sémantique du terme « data » confèrent ainsi à la donnée un caractère brut qu'il ne serait pas nécessaire de déconstruire. Cette signification persiste aujourd'hui dans la distinction, souvent avancée en science de l'information, entre les termes « donnée », « information » et « connaissance »³. La relation entre ces notions est régulièrement présentée sous la forme d'une pyramide de la connaissance qui les organise hiérarchiquement. La donnée préexiste, puis l'information et enfin la connaissance, le passage d'un niveau à l'autre s'effectuant par des opérations de « distillation »⁴.

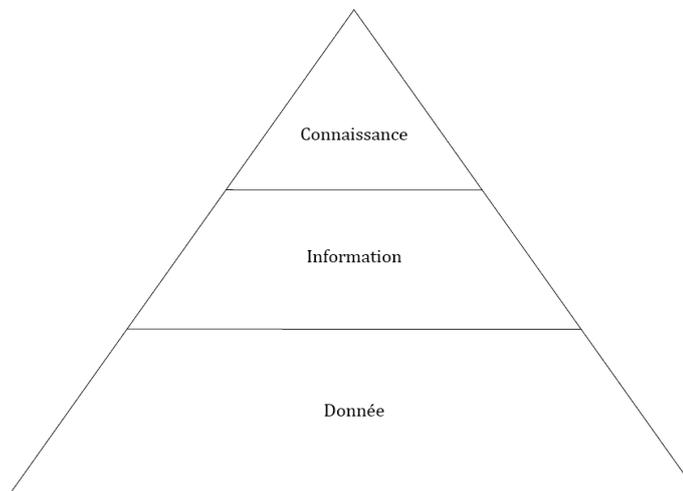


Illustration 1 : Pyramide des connaissances

En théorie de l'information, la *donnée* est une description élémentaire d'une réalité qu'elle objective. Codée sous forme numérique, la donnée est un élément brut, auquel aucun contexte n'est associé. Par exemple, la mesure de l'occupation d'une voirie est une donnée.

¹ Daniel Rosenberg, « Data before the fact » dans Lisa Gitelman (ed.), *'Raw Data' is an Oxymoron*, Cambridge Mass., MIT Press, 2013, p. 15-40.

² *Ibid.*

³ Il faudrait aujourd'hui ajouter à cette pyramide une quatrième catégorie en amont de la donnée : la « trace », c'est-à-dire l'ensemble des signaux captés, dont certains ne sont jamais mis en donnée et dont l'usage ultérieur n'est pas toujours préformaté lors de leur production.

⁴ Selon David Weinberger, *“Information is to data what wine is to the vineyard: the delicious extract and distillate”* David Weinberger, *Too Big to Know*, New York, Basic Books, 2011, p. 2 ; cité par Rob Kitchin, *The data revolution: Big data, open data, data infrastructures and their consequences*, s.l., SAGE publications, 2014, p. 10.

Elle est indépendante de son créateur et peut aisément circuler et être utilisée dans différents contextes.

L'*information* est une donnée qui acquiert du sens. Etymologiquement, information provient du latin *in forma*. L'information se caractérise par une mise en forme des données. Pour poursuivre notre exemple, la traduction de la donnée d'occupation de la voirie dans un code à trois couleurs indiquant l'état de circulation sur chacun des tronçons est une information. Une même donnée peut donner lieu à des informations de formes différentes. L'information a un contenu sémantique plus riche que les données. En étant contextualisée, l'information est davantage marquée par la subjectivité de son producteur.

Enfin, la *connaissance* est une interprétation des informations¹. Elle est un système de pensée qui relie de manière cohérente des informations pour révéler une image plus large d'un phénomène. Elle implique une opération cognitive (organiser, synthétiser, associer, analyser, raisonner, etc.) et est, par nature, personnelle et subjective puisqu'elle est internalisée par la personne qui la formule ou la reçoit. Par exemple, le fait que le réseau routier de l'agglomération lyonnaise est fortement congestionné les lundis à 18h est une connaissance.

Comme tout « investissement de forme »², le passage de la donnée à l'information puis à la connaissance peut être caractérisé sous la forme d'une dialectique de la perte et du gain³. Au passage d'une catégorie à une autre, ce que la donnée gagne en signification, elle le perd en objectivation. En acquérant une richesse sémantique conquise par les opérations successives de mise en forme et d'interprétation, la donnée n'est plus un objet stable qui peut aisément circuler d'une situation à une autre et être mis en relation.

Distinguer la donnée des faits, de l'information et de la connaissance révèle les propriétés qui lui sont couramment associées. La donnée est un élément brut décrivant une réalité. Elle a un caractère objectif qu'il n'est pas nécessaire de déconstruire puisque la donnée est *donnée*.

¹ Il faut cependant noter que la connaissance n'est pas toujours associée à des données ou des informations, elle peut être tacite, issue de croyances ou d'opinions.

² Laurent Thévenot, « Les investissements de forme » dans Laurent Thévenot (ed.), *Conventions économiques*, Paris, PUF, 1986, p. 21-71.

³ Bruno Latour, « Les “vues” de l'esprit », *Culture technique*, 1985, n° 14, p. 4-30 ; Bruno Latour, « Le topofil de Boa Vista ou la référence scientifique -montage photo-philosophique », *Raison Pratique*, 1993, n° 4, p. 187-216.

Ces détours par les sciences de l'information et la science informatique dessinent une première définition de la donnée qui s'inscrit dans une épistémologie de la mesure proche des sciences de la nature. Selon ce modèle métrologique réaliste, la donnée est objective et représente de manière fidèle une réalité qui lui préexiste. Elle se manifeste dans la quête à l'exactitude des producteurs de données et dans le souci de produire une donnée toujours plus fine et précise. La qualité d'une donnée se mesurera au degré d'écart entre la représentation et la réalité. Si la donnée ne correspond pas à la réalité perçue, c'est que la mesure est fautive.

Dans cette optique, la donnée est le résultat d'une mesure qui est collectée par des dispositifs sociotechniques variés qui ne cessent de se perfectionner. Les données géographiques étaient produites par des géomètres allant sur le terrain pour effectuer des relevés géodésiques, aidés par des « portes mire », qui tenaient le piquet pendant que leurs collègues effectuaient les mesures via un tachéomètre. Les mesures étaient ensuite recalculées manuellement avec des marges d'erreur plus ou moins grande. Enfin, elles étaient fournies à des dessinateurs, qui, armés de crayons, de pinceaux et d'aquarelles, dessinaient les cartes. Aujourd'hui, ces cartes sont produites à partir de prises de vue aérienne, dont la résolution ne cesse d'augmenter. Ces photographies sont ensuite corrigées par de complexes calculs qui redressent leur géométrie afin de les faire correspondre à une carte plane. Une évolution similaire pourrait être narrée pour la mesure des informations de trafic routier : aux comptages manuels se sont succédés les boucles à induction intégrées dans la chaussée, puis les capteurs Bluetooth et les *Floating Car Data* transmises par le réseau cellulaire.

Cette courte présentation souligne la diversité des dispositifs utilisés dans cette quête d'une mesure toujours plus précise d'une réalité. Ce schéma linéaire ne résiste cependant pas à l'observation des pratiques de production de la donnée. En effet, si les campagnes d'orthophotographie sont des outils essentiels à la production de données géographiques, les relevés topographiques n'ont pas disparu et sont toujours utilisés pour dessiner des plans. De même, les nouvelles générations de capteurs n'ont pas remplacé les comptages humains des véhicules, réalisés régulièrement par des étudiants postés à des carrefours aux heures de pointe. Pourquoi maintenir une diversité des techniques de mesure alors que les technologies s'améliorent ? Parce que ces différentes techniques ne mesurent pas la même chose et ne répondent pas aux mêmes besoins. Ce premier questionnement nous invite à interroger le caractère réaliste des données. Qu'est-ce que l'on mesure ? Pourquoi le mesure-t-on ?

Pour répondre à ces questions, à la pyramide de connaissance précédemment présentée, il nous faut ajouter un quatrième élément : le monde. Le « monde » est tout ce qu'il advient ou pourrait advenir. Le monde se caractérise par son immanence et sa nature incertaine et instable. Luc Boltanski l'oppose à la réalité qu'il définit comme l'ensemble des mises en forme et des représentations du monde¹. La réalité permet aux individus de s'accorder sur un état des choses et de construire du commun. L'enjeu n'est pas d'opposer monde et réalité, mais de comprendre le passage de l'un à l'autre.

Ce passage du monde à la réalité est l'objet de ce chapitre. Je propose pour cela d'explorer les opérations concrètes de *mise en données* de l'espace urbain. Dans la continuité de la nouvelle sociologie des sciences développée au sein de l'École des Mines² et prolongée par Alain Desrosières³, ce récit ne s'arrête pas aux opérations de *mise en données*, mais celles-ci seront illustrées par l'étude des usages administratifs de ces données. Cette approche combine une analyse *internaliste* de la donnée, qui s'intéresse aux caractéristiques techniques de l'outil, au formalisme et au codage des éléments représentés, et une sociologie *externaliste* attentive à la circulation et à l'usage de cet instrument. Comme nous le verrons, il est indispensable de penser ensemble la fabrique et l'usage des données pour en saisir tous les enjeux sociopolitiques. Cette proposition d'une sociologie de la donnée urbaine permet d'être attentif tout autant à ce que *sont* qu'à ce que *font* les données. Ces deux questions constituent la trame de fond de ce chapitre : quelles sont les principales caractéristiques des données ? Qu'est-ce que la mise en données fait à l'espace urbain ?

Pour répondre à ces questions, je m'appuie sur la notion d'attachement conceptualisée notamment par Bruno Latour et Antoine Hennion, qui me permet de questionner les propriétés les plus couramment admises des données. D'une part, l'idée d'une autonomie des données, entités immatérielles que l'on pourrait aisément mettre en circulation, ne résiste pas à l'étude empirique de la *mise en données*. Celle-ci met en évidence l'ensemble indémêlable de liens, associant pêle-mêle des acteurs, des dispositifs techniques, des conventions de

¹ Luc Boltanski, *De la critique : Précis de sociologie de l'émancipation*, Paris, Gallimard, 2009, p. 94.

² Madeleine Akrich, Michel Callon et Bruno Latour, *Sociologie de la traduction : Textes fondateurs*, s.l., Presses de l'École des Mines, 2006 ;

Pour une application de cette démarche à l'étude de la production des faits scientifiques, voir par exemple Bruno Latour, *Pasteur : guerre et paix des microbes : Suivi de Irréductions*, Paris, La Découverte, 1984, 363 p.

³ Alain Desrosières, *La politique des grands nombres. Histoire de la raison statistique.*, Paris, La Découverte, 1993.

mesure, etc., qui fait exister la donnée. Ces attachements sont constitutifs de la donnée. Dès lors, pour mettre en circulation des données, il ne faut pas opposer attachement et détachement, mais être attentif aux attachements et faire le tri entre les bons et les mauvais attachements. Pour paraphraser Bruno Latour, la question n'est pas de savoir si les données sont autonomes ou attachées, mais de savoir si elles sont bien ou mal attachées¹. Dès lors, si ce chapitre ne traite pas explicitement d'*open data*, il est crucial pour la compréhension de la suite de cette thèse. Pour mettre en circulation les données, il est nécessaire de partir des liens qui les attachent, de les éprouver pour rendre possible le détachement de certains et l'attachement à d'autres.

D'autre part, tout autant qu'elle est attachée, la donnée attache. Elle produit des liens entre la ville d'un côté et les acteurs qui agissent sur celle-ci de l'autre. Cependant, en reliant ces mondes, à l'instar de tout attachement, la donnée introduit une différence dans un cours d'action : elle ne reflète pas une réalité qui serait déjà là, mais elle la fait exister. Cette caractéristique, qui souligne le pouvoir d'agir de la donnée, son *agency*, questionne le caractère réaliste souvent associé aux données. Les données ne sont pas un intermédiaire entre deux mondes déjà-là, mais une médiation qui façonne la ville tout autant que les collectifs qui s'y attachent. Comme le soulignent Emilie Gomart et Antoine Hennion², toute la force de la notion d'attachement est de ne pas se restreindre à décrire un réseau d'attachements, mais également de prendre en compte ce qu'il advient, ce qui émerge de la rencontre entre ces mondes.

En s'appuyant sur la notion d'attachement, ce chapitre étudie le « faire » des données, c'est-à-dire « à la fois le fait qu'elles sont faites et le fait qu'elles font faire »³. Pour cela, notre analyse des opérations de mise en données de l'espace urbain débute par la description des données de circulation routière qui met en évidence le fait que les données sont constituées d'attachements formant un fragile réseau sociotechnique. Dans une seconde partie, en mettant en évidence les conventions sur lesquelles reposent les données, nous verrons que les données attachent des « métiers » à l'espace urbain au travers d'une

¹ Bruno Latour, « Factures/Fractures. De la notion de réseau à celle d'attachement » dans André Micoud et Michel Peroni (eds.), *Ce qui nous relie*, La Tour d'Aigues, Editions de l'Aube, 2000, p. 189-208.

² Emilie Gomart et Antoine Hennion, « A Sociology of Attachment: Music Lovers, Drug Addicts » dans John Law et John Hassard (eds.), *Actor Network Theory and After*, Oxford/Malden MA, Blackwell Publishers, 1999, .

³ Antoine Hennion, « D'une sociologie de la médiation à une pragmatique des attachements. », *SociologieS*, 2013.

représentation spécifique de celui-ci. Enfin, une troisième partie est consacrée aux données de référence du système d'information géographique de l'agglomération. En étudiant leurs propriétés, est mis en évidence la spécificité de l'attachement qu'elles produisent contribuant à donner consistance parallèlement à l'institution communautaire et à son territoire. En conclusion, nous reviendrons sur les caractéristiques des données métiers et des données de référence qui produisent des attachements différents entre la ville et les collectifs qui s'y attachent, mettant en évidence deux formes spécifiques d'agencement.

Avant d'entamer notre exploration, un bref rappel historique s'impose. Si le terme de données est antérieur à l'apparition de l'informatique, il n'est pas usité au sein de la communauté urbaine de Lyon. Il faut attendre le début des années 80 et l'apparition des premières bases de données relationnelles au sein du service de l'assainissement¹ pour que celui-ci fasse son apparition. La « donnée » est indissociable de l'infrastructure informatique et de la numérisation des représentations de l'espace urbain.

I. Les données sont faites d'attachements

A partir de l'exemple des informations de circulation routière produites par la Métropole de Lyon, la première partie de ce chapitre tente de caractériser ce qu'est une donnée par l'analyse des opérations de *mise en données* de l'espace urbain. Au travers de la description de la chaîne de production de ces informations, nous verrons qu'il n'existe rien de tel qu'une « donnée » qui apparaîtrait de manière évidente à l'observateur et que l'on pourrait facilement isoler. Ni immatérielles, ni autonomes, les données sont constituées de multiples attachements. Elles sont le résultat d'un tissu de relations liant des entités hétérogènes. Ces attachements qui font la donnée sont toujours fragiles et nécessitent un travail quotidien pour en assurer la solidité.

¹ Le service de l'assainissement est un pionnier dans l'utilisation de l'informatique à la Courly qu'il substitue progressivement à des métiers bien établis (cartographe, dessinateur, etc.). Cette évolution s'est effectuée en plusieurs étapes avec tout d'abord la migration de la production de plans ou de cartes de la planche à dessin à la table à digitaliser, puis à l'écran avec le développement des logiciels de dessins et de cartographies numérisés (CAO/DAO). Après avoir numérisé l'ensemble de ses plans, le service de l'assainissement va constituer une base de données de son réseau afin d'en faciliter les mises à jour et de permettre des modélisations. Apparaissent alors les premiers systèmes de gestion de base de données (SGBD) visant à organiser la gestion et le stockage de ces données numérisées et à automatiser les opérations de mise à jour.

A. Du signal au centre de calcul : la chaîne de la « donnée » de circulation routière

La donnée de circulation routière est insérée au sein d'une longue chaîne sociotechnique. Du signal capté sous la chaussée au « centre de calcul » pilotant à distance les feux de circulation, la « donnée » change d'état tout au long du processus de *mise en données* de la congestion routière. Ce cas met en avant l'impossibilité de penser la donnée indépendamment de tous les attachements qui la constituent.

La première étape de cette chaîne de la donnée est le capteur. Pour mesurer la circulation routière, la communauté urbaine de Lyon dispose d'environ 760 postes de comptages du trafic automobile et cycliste. Ces capteurs sont majoritairement des boucles à induction, placées sous la chaussée, qui détectent le passage ou la présence d'un véhicule.



Illustration 2 : Boucles à induction détectant le passage ou la présence d'un véhicule sur la chaussée

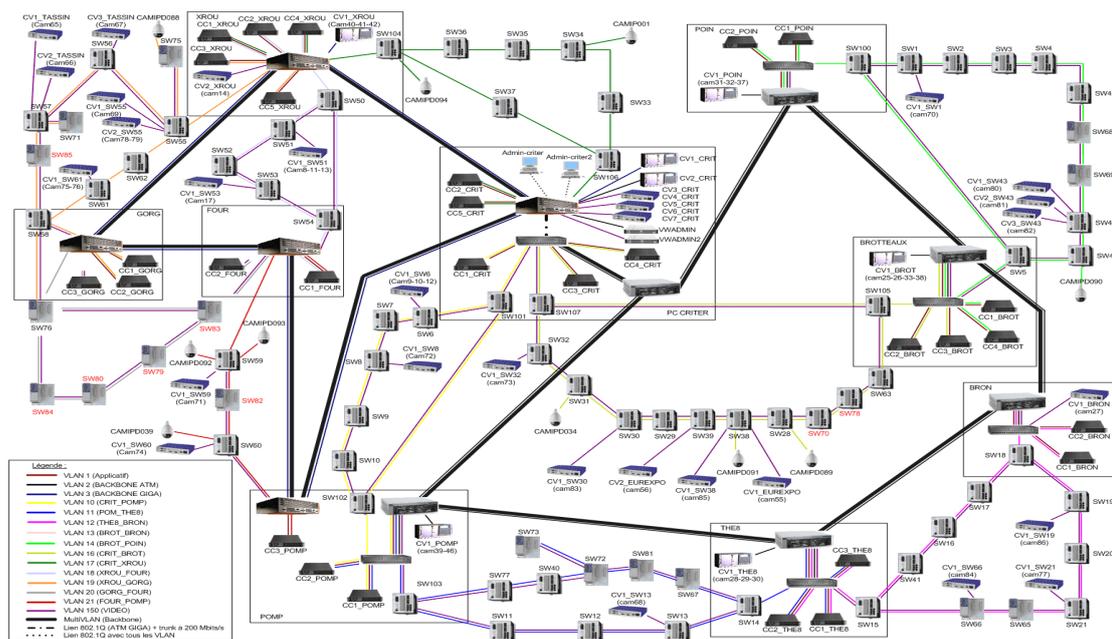


Illustration 3 : Plan du réseau de télécommunications de régulation des feux de circulation

Ces capteurs étalonnés produisent un signal qui est transmis à un contrôleur de feux de circulation relié en fibre optique à un PC Central¹. Au centre de ce réseau de télécommunications, le « PC Criter » est un gros « calculateur » qui centralise et agrège les signaux des 760 postes de comptage pour produire une mesure du trafic routier. Ces mesures deviennent ensuite données en étant intégrées au sein d'une base de données. On peut alors obtenir, en temps réel, un état de circulation sur l'agglomération. Puis, ces données deviennent information lorsqu'elles sont traduites en débit (le nombre de véhicules par unité de temps) ou en taux d'occupation (taux de temps d'occupation d'une boucle par unité de temps).

Des boucles à induction, on tire des débits [le nombre de véhicules par minute, par six minutes, ou même par heure], et des taux d'occupation, c'est-à-dire pendant combien de temps la voiture reste sur cette boucle. Et cette valeur nous sert à caractériser la congestion en fonction de seuils. Pour la diffusion auprès du public, on décrit ensuite si c'est rouge, orange ou vert, ils appellent ça

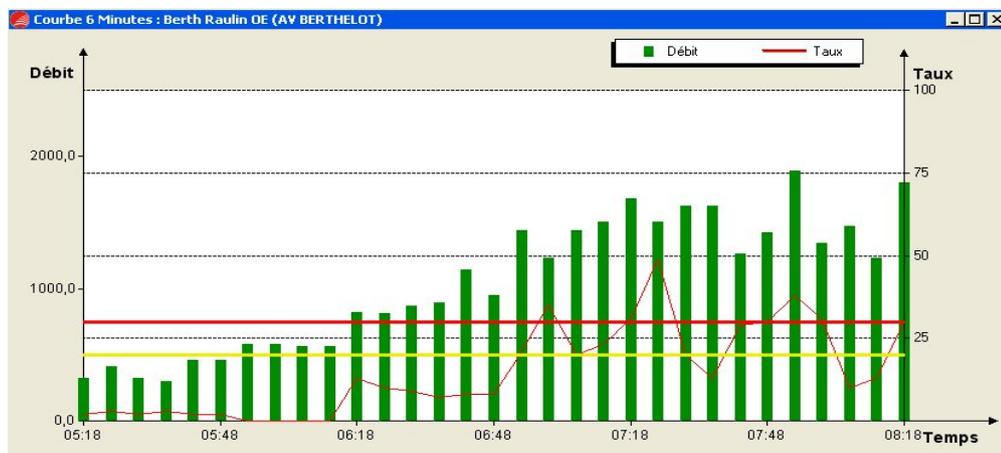
¹ « Ces contrôleurs sont reliés en fibre optique au PC, c'est un réseau maillé, c'est une sorte de pseudo internet pour les feux de régulation. » (Source : entretien avec Pierre, responsable technique, PC Criter, Grand Lyon (9 mars 2014))

La direction de la voirie a en effet construit un complexe réseau de télécommunications, spécifiquement dédié au transfert de ces données, composé de 485 kilomètres de fibres optiques, de 28 frontaux de communication, de 174 commutateurs réseaux, de serveurs industriels, etc

le trafic. On réunit toutes ces données dans nos bases, c'est utilisé par les opérateurs de régulation et d'information¹.

Mesures : Courm N/S (QU GAILLETON)					
Horodate : 11/03/2009,08:25:00					
	Voie 1	Voie 2	1 Mn lissée	1 Mn	6 Mn
Débit	960	1020	1620	1980	1570
Taux	54	42	51	48	29
Vitesse					

Illustration 4 : Mesures d'une boucle de comptage (débit et taux à la minute)



¹ Source : entretien avec Pierre, responsable technique, PC Criter, Grand Lyon (9 mars 2014)

Illustration 5 : Mesures d'une boucle de comptage (évolution des débits et des taux sur une période de six minutes)

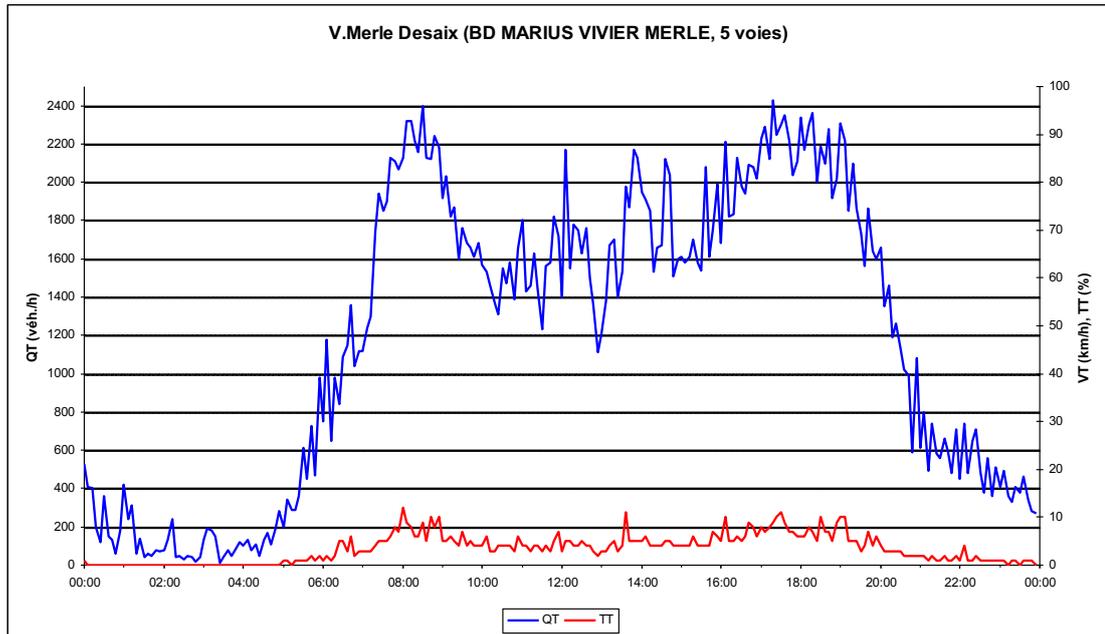


Illustration 6 : Mesures d'une boucle (débit et vitesse moyenne) sur une plage de 24h

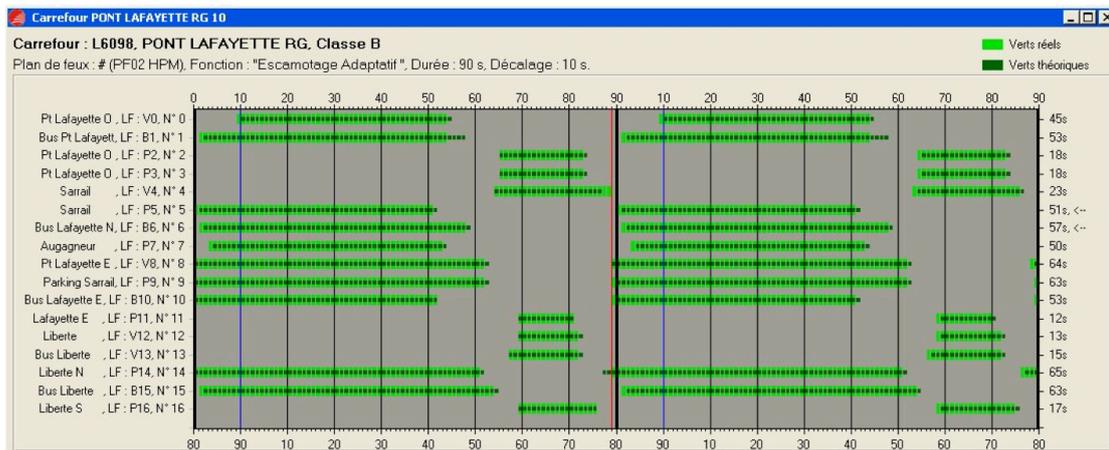
Au sein du PC Criter, ces informations de débits et d'occupation sont ensuite à nouveau traduites par les « traficiens », des experts des flux de déplacements, en plans de gestion du trafic urbain. Ces plans sont mis en œuvre par des opérateurs en charge de la régulation du réseau routier¹, réunis dans un poste de contrôle où ils font face à une dizaine d'écrans, retransmettant des images de vidéosurveillance, encadrant une carte des principaux axes de circulation de l'agglomération.

¹ Ces opérateurs se relaient sur deux périodes quotidiennes de huit heures et une astreinte est organisée chaque nuit.



Illustration 7 : Opérateurs du PC Criter de la Métropole de Lyon

Pour assurer la régulation du réseau routier, ces opérateurs adaptent les plans de feux aux conditions de circulation. Les plans de feux sont un ensemble de cycles de fonctionnement du carrefour lors desquels les différentes phases de vert s'enchaînent. Ces séquences, appelées « diagramme de vert », varient selon la géométrie du carrefour et les intensités du trafic. Ils sont programmés afin d'obtenir un écoulement optimal des véhicules selon un système d' « ondes vertes ».



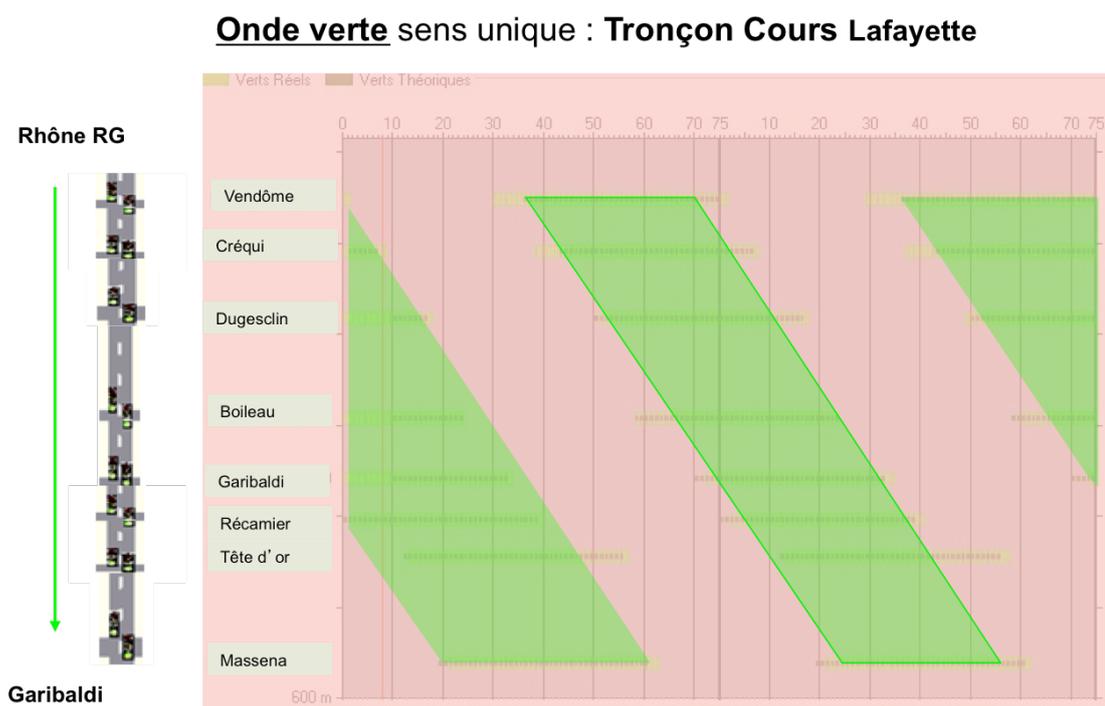


Illustration 9 : « Ondes vertes » : programmation d'un passage au vert successif des feux du cours Lafayette

En fonction des conditions de circulation relevées par les boucles à induction, différents plans de feux sont utilisés pour réguler les flux de trafic. Le processus de décision des plans de feux à mettre en œuvre repose essentiellement sur de complexes algorithmes de priorité qui déterminent automatiquement le plan de feu à appliquer, selon les conditions de trafic, le calendrier, l'heure ou des événements programmés (travaux, manifestations, etc.)¹.

Par exemple, il y a un événement au tunnel de Fourvière, il faut le fermer. Tout de suite, il y a un scénario qui se met en place tout seul, où, sur une zone définie, tous les plans de feux sont modifiés pour favoriser l'axe secondaire et éviter une trop forte congestion. Il y a plusieurs scénarios qui sont déjà définis en amont et qui sont priorisés entre eux, en fonction des événements (coupure ou réduction de voies). [...] La difficulté c'est lorsque des événements multiples se produisent et que les scénarios sont contradictoires : l'un dit « il faut prioriser cet axe là », l'autre dit le contraire. Donc là, il faut savoir quel plan de feu mettre au bon moment en fonction de deux événements majeurs².

Une fois déterminés au sein du PC Criter, ces plans sont transmis à un « réseau d'actionneurs » qui permet de piloter à distance les carrefours à feux. Chaque carrefour à feux

¹ De manière ponctuelle, les opérateurs prennent directement en main la détermination des plans de feux à appliquer. Au total, 3 millions de commandes de changements de plans de feux pour réguler le trafic sont effectuées chaque année (dont environ 5000 directement par l'opérateur).

² Source : entretien avec Pierre, responsable technique, PC Criter, Grand Lyon (9 mars 2014)

est relié à un « contrôleur », « un ordinateur, situé dans une boîte au pied du feu. Toutes les boucles de comptage sont connectées à ce contrôleur, tous les feux sont connectés à ce contrôleur, et c'est à l'intérieur de ce contrôleur qu'il y a les calendriers et les plans de feux¹. » Les contrôleurs de carrefour prennent la forme d'une armoire située en bordure de chaque carrefour de circulation. Ce sont des ordinateurs spécialisés dédiés à la gestion des feux de circulation en régulation de trafic. Ces automates programmables permettent aux régulateurs d'agir à distance sur ces feux de circulation en faisant varier le temps de vert en fonction des conditions de circulation. « On fonctionne avec des plans de feux, souvent, c'est pics du matin et du soir, et journée. Le matin, on favorise l'entrée dans Lyon, le soir, la sortie de Lyon². » L'enjeu est de fluidifier le trafic sur l'agglomération.



Illustration 10 : Contrôleur de carrefour à feux

Ainsi, l'énoncé qui circule le long de la chaîne d'information routière est traduit lors de son association successive à une nouvelle matérialité. Le signal capté par la boucle à induction devient mesure lorsqu'il est agrégé à d'autres signaux par les calculateurs du PC Criter. Ces mesures deviennent données en étant insérées dans des bases de données. Les données se transforment en information lorsque l'on produit des taux comparables et

¹ Source : entretien avec Pierre, responsable technique, PC Criter, Grand Lyon (9 mars 2014)

² Source : entretien avec Pierre, responsable technique, PC Criter, Grand Lyon (9 mars 2014)

visualisables. Ces informations deviennent enfin des plans stratégiques quand les « traficiens » les croisent avec les plans de feux pour répondre aux objectifs de régulation du réseau.

Au cœur de cette chaîne sociotechnique, le PC Criter permet à la fois de suivre en continu la circulation routière sur l'agglomération lyonnaise et de réguler ces flux de circulation en pilotant à distance les feux de circulation. Il constitue un véritable « centre de calcul¹ » au sens où l'entend Bruno Latour. Il rassemble en un même espace les mesures provenant des boucles à induction, les rend compatibles et comparables et les met en relation. Cela produit une vue synoptique de l'état de la circulation routière sur l'agglomération à partir des observations faites en différents lieux. Cette activité de production de connaissance, qui s'appuie sur des « mobiles immuables² », offre de nouvelles prises pour « agir à distance » sur la ville.

La donnée d'état du trafic est ainsi insérée au sein d'un réseau d'une grande complexité : plus de 1400 feux de circulation pilotables à distance (sur les 1500 que compte l'agglomération), 764 postes de mesure du trafic, 255 caméras vidéo, 485 kilomètres de fibre optique, 28 frontaux de communication, 174 commutateurs réseaux, des serveurs industriels, des « traficiens », des équipes d'opérateurs de régulation, des algorithmes décidant des plans de feux à appliquer, etc.³. Difficile au milieu de cette complexe chaîne sociotechnique d'autonomiser la donnée. Et pour cause, loin d'être immatérielle, la donnée est constituée de ce tissu de relations liant des entités hétérogènes dont l'objectif est de réguler en temps réel la circulation routière sur les 2400 kilomètres de voirie de l'agglomération lyonnaise. Ainsi, la donnée n'existerait pas sans ces attachements qui la constituent. Ces attachements sont

¹ B. Latour, « Les “vues” de l'esprit », art cit.

² Bruno Latour définit par ce terme l'ensemble des entités (cartes, documents, etc.) à la fois mobiles, stables et combinables pris dans un processus scientifique cumulatif et qui permettent la formation d'un « centre de calcul ». Dans *La science en action*, il les désigne ainsi : « tous ces objets [qui] se situent au début et à la fin d'un cycle d'accumulation similaire ; peu importe s'ils sont proches ou éloignés, infiniment grands ou petits, infiniment vieux ou jeunes, ils finissent tous par exister à une échelle que quelques-uns peuvent dominer du regard ; à un moment ou à un autre, ils prendront tous la forme d'une feuille de papier qui peut être archivée, épinglée au mur et combinée à d'autres » (p. 545).

B. Latour, *La science en action*, op. cit.

³ Source : Chiffres de 2014, http://www.onlymoov.com/decouvrir/questions-reponses/faq/c189b8ddfc69157918a4d101a42cb91c/?tx_glfq_faq%5Bfaq%5D=9&tx_glfq_faq%5Baction%5D=show&tx_glfq_faq%5Bcontroller%5D=Faq#.VnvcppPhBE4 (Consulté le 24 décembre 2015)

toutefois fragiles : des opérations quotidiennes de maintenance sont nécessaires pour assurer la solidité des associations qui lient chacune des entités de ce réseau.

B. Des attachements fragiles quotidiennement solidifiés

La donnée d'état de la circulation routière est insérée au sein d'une longue chaîne sociotechnique, du capteur de mesure inséré sous la chaussée au pilotage à distance des feux de circulation. Toutefois, les associations qui lient les entités de ce réseau peuvent rendre impossible la bonne régulation du réseau routier. Les facteurs de dissociation de ces entités sont divers : une boucle de comptage défectueuse, une défaillance du réseau de télécommunication, une panne du contrôleur de feux, une mauvaise interprétation de l'opérateur, un mauvais plan de feux appliqué, etc. Par exemple, les boucles à induction ont une durée de vie limitée et leur emplacement sous la chaussée les rend vulnérables aux chantiers réalisés sur ces voiries.

Le problème avec les boucles analogiques, c'est qu'à chaque fois qu'il y a un véhicule qui passe, ça s'ouvre et ça se referme. Ce relais, on a beau prendre un relais de qualité, il a une durée de vie limitée. Le problème, c'est qu'en fin de vie, le relais ne lâche pas d'un coup. Il va osciller par exemple, et il nous pourrit la mesure. Voilà. Aujourd'hui, on sait détecter quand il est complètement collé. La phase où il va rebondir trois fois, ce qui va nous faire croire qu'il y a trois véhicules qui sont passés alors qu'ils ne sont pas passés, elle est beaucoup plus difficile à détecter. [...] Tout ça conjugué fait que la continuité d'une donnée de qualité avec des boucles inductives est compliquée. Un détecteur qui va être en fin de vie, il va rebondir trois fois, on va pas le détecter tout de suite, c'est quand on va analyser les données, mais on ne le fait pas tous les jours, on va dire : « *il y a eu une augmentation de trafic là ? Bizarre* ». Après, généralement, il passe par cette phase là, et un jour il colle, alors dès qu'il est collé, le lendemain matin, on sait. C'est la phase intermédiaire. En comptage, cette phase là, elle est préjudiciable¹.

Pour maintenir « la continuité d'une donnée de qualité », le réseau de capteurs est soumis à une vigilance toute particulière. Des analyses sont effectuées a posteriori afin de déceler certaines mesures « bizarres » et d'identifier les capteurs défectueux. La solidité du réseau de la donnée d'état de trafic est assurée grâce à ces opérations de vérifications régulières du bon fonctionnement du système et des équipements.

Après, il y a une deuxième phase qui est le débriefing, vérifier si ça a bien marché. Il faut bien voir que c'est toute une chaîne technique. Il peut y avoir un problème au carrefour, il peut y avoir un problème réseau, il y a des fois, c'est des problèmes exogènes d'ailleurs. Aujourd'hui, quand on a une bonne supervision de l'ensemble, quand on s'aperçoit qu'à un endroit, avec ce que l'on avait mis comme plan de feu, on aurait du évacuer 800 véhicules / heure par voie, et la boucle nous dit en fait, c'était 300, on se pose la question : pourquoi ? On va d'abord vérifier qu'on avait bien mis ce qu'il fallait sur le contrôleur de feu. Maintenant, généralement c'est bon, on est bien rodé. Il peut arriver dès fois qu'il y a eu un petit problème de programmation, mais c'est quand même

¹ Source : entretien avec Bruno, responsable du PC Criter, Grand Lyon (20 mars 2015)

relativement rare. Parfois, c'est aussi une panne d'une boucle qui fait que l'on n'a pas ce qu'il faut. Ça nous permet d'identifier les pannes. On traite toutes ces informations pour avoir la ressource disponible qui fonctionne bien au bon endroit. Une fois qu'on a vérifié que l'on a bien mis le plan de vert, c'est qu'il se passe autre chose¹.

Lorsqu'une panne d'un équipement du réseau est repérée, le pôle technique du PC Criter prend le relais. Cette équipe est en charge de la maintenance préventive et curative du réseau. Elle assure le remplacement des éléments défectueux de la chaîne de la donnée et règle les capteurs de comptage. Ainsi, la consistance des assemblages sociotechniques est mise à l'épreuve dans différentes situations. Leur robustesse, leur irréversibilité, leur autonomie ne sont jamais acquises pour toujours². Elles ne reposent que sur le travail continu d'attachement, d'alliances nouées, pour allonger le réseau sociotechnique et assurer sa solidité. Souvent peu visible, le travail quotidien des « petites mains³ » de la chaîne de la donnée est indispensable pour renforcer la robustesse de ces attachements. Le travail de l'équipe du PC Criter se résume ainsi à la production et l'entretien d'un vaste réseau sociotechnique, qui se déploie de la description de l'espace urbain à l'action sur celui-ci.

Encadré 2 : Les limites de la quantification

Ces opérations d'entretien de la chaîne de la donnée ne sont pas toujours suffisantes pour assurer le travail de régulation du réseau routier. Parfois, « *il se passe autre chose* » : malgré tout l'investissement réalisé dans la définition et la mise en œuvre des plans de régulation optimaux à partir des données de circulation, des facteurs exogènes viennent contrarier le travail patiemment construit par les opérateurs à leur grand désarroi.

Notre ennemi aujourd'hui, ce sont les gens qui se garent en double file. Les livreurs, les machins, les trucs à l'heure de pointe, ça fout en l'air notre système. Sur l'avenue Berthelot, un déménageur qui va se garer

¹ Source : entretien avec Bruno, responsable du PC Criter, Grand Lyon (20 mars 2015)

² Dans son enquête sur l'échec du projet de métro automatique « Aramis », Bruno Latour met en avant tout le travail nécessaire et toujours à refaire pour rendre solide et pérenne un projet technique. « *Aussi puissant, aussi irréversible et indispensable, et donc aussi réel que soit un moyen de transport, on peut toujours le déréaliser en partie [...]. Ces énormes monstres techniques vieux de cent ans ne sont pas plus réels qu'Aramis, vieux de quatre ans, n'est irréel : à tous, il leur faut des alliés, des amis, des longues chaînes de traducteurs. Il n'y a pas d'inertie, il n'y a pas d'irréversibilité, il n'y a pas d'autonomie pour les tenir en vie. Derrière ces trois mots de la philosophie des techniques [...], il y a le travail continu pour attacher ou détacher des locomotives et des wagons, des élus locaux et des ingénieurs, des grévistes et des usagers, des parcours et des flux de trafic* ».

Bruno Latour, *Aramis ou l'amour des techniques*, Paris, La Découverte, 1992, p. 78-79.

³ L'adjectif « petit » ne fait pas référence à l'importance de ce travail, qui est crucial, mais à sa faible visibilité et le peu de reconnaissance dont il est l'objet dans la chaîne de production de l'information. Cf. Jérôme Denis et David Pontille, « Travailleurs de l'écrit, matières de l'information », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 1-20.

en double file, et qui déménage à l'heure de pointe entre 7h30 et 9h30 qui va immobiliser une voie, ça fout notre boulot en l'air. Ça nous énerve. C'est agaçant. On met en place, on travaille à faire en sorte que ça marche bien, on optimise, on a des outils, on fait un gros travail, et bim, un matin donné, on va voir que Berthelot va redevenir tout rouge, voilà, on a envie d'aller lui foutre une prune. Il y a des endroits aussi où certains bus, pour des raisons de sécurité que je comprends, immobilisent une voie régulièrement. La fréquence de la ligne monte, ça fait que nous le feu est vert, mais comme le bus est sur la voie de droite, les véhicules arrivent derrière, déboitent, ce qui gêne l'autre file, donc au lieu d'avoir deux fois une voie, deux fois 1 en débit sur chacune des voies, on se retrouve globalement avec un système qui a un débit pour les véhicules de 0,7, parce que comme les voitures changent de voie, le débit ralentit. On a plein de trucs comme ça, donc ça nous permet aussi d'avoir une idée des facteurs exogènes¹.

Ces événements soulignent que la ville n'est pas réductible à un modèle quantifié. Le camion de déménagement garé en double file, le véhicule en panne sur la chaussée ou encore les véhicules dépassant un bus qui stationne longuement sont autant d'événements du monde venant contester la réalité de la circulation routière construite par la mise en données. S'ils ne disposaient pas d'un dispositif de caméras de vidéosurveillance, les régulateurs du réseau n'auraient aucune connaissance de ces facteurs exogènes qui entravent le bon écoulement des flux et ne pourraient plus agir sur la circulation routière.

C. Conclusion : le réseau sociotechnique de la donnée

Ni autonomes, ni immatérielles, les données reposent sur de multiples attachements hétérogènes à des personnes, des pratiques, des technologies, des institutions qui les produisent, les maintiennent, les mettent en circulation et les utilisent. Elles sont indissociables de l'ensemble des entités qui leur sont associées : les capteurs de mesure, les ordinateurs, les équipes de maintenance, les régulateurs du réseau, les automobiles, les feux de circulation, etc. Le réseau sociotechnique de la donnée lie tous ces éléments, à la fois humains et non humains². Les données ne remplissent leur mission que parce qu'elles sont insérées dans un réseau, composé de multiples dispositifs, qui les produisent et autorisent leurs usages. Elles ne tiennent que par l'ensemble des connexions qu'elles établissent. Des longues chaînes de petites actions qui font, composent, produisent une totalité : ici l'état de la circulation routière, ailleurs l'offre de transport en commun sur l'agglomération ou les circuits de collecte des déchets ménagers. Mais, cette chaîne est fragile. Il suffit qu'un véhicule

¹ Source : entretien avec Bruno, responsable du PC Criter, Grand Lyon (20 mars 2015)

² Michel Callon, John Law et Arie Rip (eds.), *Mapping the Dynamics of Science and Technology : Sociology of Science in the Real World*, London, Mac Millan, 1986, 242 p ; B. Latour, *La science en action*, *op. cit.*

stationne trop longuement sur une boucle de comptage, et la donnée perd de sa fiabilité. Elle ne représente plus la congestion, juste un livreur un peu trop lent, et c'est toute la chaîne qui s'écroule. La donnée ne répond plus à l'usage. Le travail de totalisation est ainsi toujours à reprendre et ne s'exerce que sous l'œil vigilant des opérateurs. Un travail quotidien, toujours répété, de maintenance, d'entretien et d'allongement de la chaîne des données est nécessaire pour assurer la solidité de la donnée. C'est à ce prix que l'*agency* du réseau¹, la capacité de la donnée à recomposer et transformer l'espace urbain en le représentant, se réalise. Car, si la donnée est constituée d'attachements, elle est elle-même constitutive d'attachements. Or, la force des liens fait faire quelque chose. La partie suivante de ce chapitre est précisément consacrée à l'agentivité de la donnée (son *agency*²), c'est-à-dire sa capacité à introduire une différence dans un cours d'action.

II. Représenter l'espace urbain : conventions et usages des données

Après avoir mis en évidence les attachements constitutifs des données, la seconde partie de ce chapitre est consacrée aux modalités de la donnée comme attachement. Il s'agit de se demander : qu'est-ce que la donnée attache ? Et comment attache-t-elle ? Pour cela, à partir de la présentation de quelques exemples de données peu « réalistes », j'analyse les conventions sur lesquelles reposent les données. Ces conventions, qui restreignent la représentation de l'espace urbain à certains éléments, sont déterminées par les usages souhaités de la donnée. De cet attachement de la donnée à des situations et des finalités particulières découle une forme d'agencement territorial spécifique associant une donnée, un « métier » et une représentation de l'espace.

A. Des données peu « réalistes » : représenter l'espace urbain pour agir dessus

L'épistémologie réaliste, selon laquelle la ville préexisterait aux données qui ne feraient que la refléter, ne résiste pas à la mise en évidence des conventions sur lesquelles reposent les

¹ Michel Callon, « Sociologie de l'acteur réseau » dans Madeleine Akrich, Michel Callon et Bruno Latour (eds.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'Ecole des Mines, 2006, p. 267-276.

² La notion d'*agency* est au cœur de la théorie de l'acteur-réseau. Elle signifie qu'un dispositif dispose d'une capacité d'action par l'ensemble des relations qui le constitue.

données. Les représentations de l'espace urbain, véhiculées par les données, sont orientées en vue d'une finalité spécifique.

1) La pluralité des représentations de l'espace urbain

Deux jeux de données distincts ont pour objet les emplacements de stationnement des deux roues. L'un est produit par la ville de Lyon, l'autre par la communauté urbaine. Ces deux données ne contiennent pas le même nombre d'entités représentées¹ ni les mêmes informations pour chacune de ces entités.

nom	numero_voi	arrondisse	longueur	gid							
RUE CHAVANNE		1	1	10	1						
identifiant	adresse	commune	mobillier2r	gestionnai	localisati	nbarceaux	capacite	anneereali	pole	observatio	gid
2R332	1, RUE CHAVANNE	Lyon 1 er	En U invers	Grand Lyon	Sur chaussée	5	10			Longueur = 10m	316

Illustration 11 : Extraits des jeux de données « Emplacement moto de la ville de Lyon » et « Parc 2RM² »

Ces différences s'expliquent par les compétences respectives de chacune des institutions. La ville de Lyon dispose d'un pouvoir de législation : elle prend les arrêtés de stationnement fixant l'emplacement des parkings pour deux roues. Mais, c'est le Grand Lyon qui installe et gère ces équipements sur l'espace public. Les modalités de production de ces données varient, ce qui se traduit par des informations différentes pour un même objet. La ville de Lyon produit la donnée à partir des arrêtés permanents de stationnement pris par la ville pour installer ces emplacements sur l'espace public de voirie. Le Grand Lyon complète cette donnée par un « relevé terrain ». Ces deux données, qui correspondent à différentes étapes de la mise en place d'un emplacement de stationnement pour deux roues, sont très hétérogènes. Et pour cause, les conventions sur lesquelles sont basées les représentations des parkings pour deux roues divergent entre les deux institutions³. Pour l'une, un parking pour deux roues est un arrêté municipal, pour l'autre, une existence concrète sur le terrain.

¹ 372 emplacements dans la donnée de la ville de Lyon contre 409 dans celle du Grand Lyon. Ces différences peuvent s'expliquer par des arrêtés non répertoriés dans le premier jeu de données ou des emplacements installés sur l'espace public sans arrêté associé.

² Extrait des jeux de données disponibles sur le site data.grandlyon.com, consulté le 17 janvier 2015.

³ Précisons qu'il n'y a pas une « vraie » et une « fausse » représentation. Les deux représentations sont justes, mais elles ne mesurent pas la même chose car les conventions préalables diffèrent.

Un même objet peut ainsi être envisagé de multiples façons. Ces représentations plurielles ne sont pas des manifestations d'une interprétation biaisée de la réalité mais elles révèlent la « richesse » de l'objet, traversé par de multiples pratiques, investi par des acteurs hétérogènes, qui en stabilisent tel ou tel aspect¹. Comme le souligne Bruno Latour : « *Ce qui fait que la multiplicité des points de vue ne vient pas d'une faiblesse de nos interprétations successives, mais bien de la richesse de l'objet lui-même. C'est parce qu'il est complexe qu'il génère autant de points de vue sur lui-même. Cette complexité est un éloge de l'objet et non pas un éloge des subjectivités qui le considéreraient de l'extérieur*² ». En fonction des réseaux sociotechniques dans lesquels sont insérés ces objets, les interprétations sur ceux-ci varient. L'espace urbain peut ainsi être représenté de façon multiple selon l'action que l'on souhaite effectuer.

De manière similaire, les multiples jeux de données d'adressage soulignent la pluralité des vues sur la ville qui coexistent. Une adresse paraît être une information simple qui ne peut souffrir d'aucune contestation : un numéro, un nom de rue, un code postal et une commune. Quatre éléments qui semblent aller de soi pour caractériser une adresse. Pourtant, l'adresse est beaucoup moins codifiée que cela (malgré les nombreuses tentatives pour normaliser l'adresse). Cette information cruciale de localisation prend en effet de multiples formes selon les usages que l'on souhaite en faire. La communauté urbaine de Lyon modélise par exemple les adresses selon trois formats : le numéro de voirie, le point de débouché et le point d'adressage. Le numéro de voirie matérialise l'implantation physique de la plaque portant l'indication du numéro dans la voie, reconnu lors d'une visite sur le terrain³. Le point de débouché est la projection orthogonale du numéro de voirie sur le filaire de voie auquel il se rapporte. Enfin, le point d'adressage est un objet symbolique ponctuel qui indique la parcelle ou le bâtiment se rattachant à un numéro de voirie. « *Parce qu'à certains endroits tu peux*

¹ Sur la multiplicité ontologique d'un objet selon les acteurs qui s'en saisissent, on pourra se référer à l'étude ethnographique des pratiques médicales réalisée par Annemarie Mol. Jérôme Denis et David Pontille ont étudié les controverses entre les contributeurs d'OpenStreetMap liées aux interprétations multiples de l'espace lors de leurs opérations de mise en données des infrastructures cyclistes.

Annemarie Mol, *The Body Multiple: Ontology in Medical Practice.*, London, Duke University Press, 2003, 216 p ; Jérôme Denis et David Pontille, « Une infrastructure évasive. Aménagements cyclables et troubles de la description dans OpenStreetMap », *Réseaux*, 2013, vol. 2, n° 178-179, p. 91-125.

² Bruno Latour, « De la politique du temps à la politique de l'espace ? » dans Elisabeth Rémy et al. (eds.), *Espaces, savoirs et incertitudes*, Paris, Ibis Press, 2005, p. 9 ; Cité par J. Denis et D. Pontille, « Une infrastructure évasive. Aménagements cyclables et troubles de la description dans OpenStreetMap », art cit.

³ Jusqu'à peu, lorsqu'il n'y avait pas de matérialisation physique ce numéro de voirie n'était pas saisi dans la base de données.

*avoir sur un porche d'entrée un numéro 28, et derrière tu as trois maisons qui sont toutes les trois adressées au numéro 28, parce qu'elles n'ont pas un numéro spécifique*¹. » Ces trois modes de représentation de l'adresse divergent encore de ceux choisis par d'autres institutions publiques ou privées. La base adresse de l'INSEE, le Répertoire des immeubles localisés (RIL), qui est essentielle pour les opérations de recensement, ne contient que les bâtiments où résident des individus. « *Donc, en gros le numéro que nous on a saisi sur le parc machin ou sur le gymnase, parce que le gymnase il a aussi une adresse, même s'il y a personne qui vit dedans, eux, ils s'en foutent. Ils ne l'ont pas dans leur adresse. Donc, le RIL, par rapport à notre base adresse, c'est pas la même chose* »². La Poste aborde les adresses sous l'angle de son métier postal : elles correspondent aux endroits où ils distribuent du courrier³. Ainsi, il y a une ambiguïté sur ce qu'est une adresse : « *la modélisation est assez compliquée parce que dès que l'on rentre, dès que l'on parle à différents métiers, les gens ne voient pas l'adresse au même endroit. Ils n'ont pas la même vue* »⁴. Dès lors, pour mettre en donnée une adresse, il est nécessaire de lever cette ambiguïté en convenant préalablement sur ce que l'on entend par l'adresse. Il faut choisir une « vue » parmi toutes celles envisageables. Est-ce qu'il faut qu'elle soit matérialisée ? Est-ce qu'il faut des résidents ? Cette opération de clarification est constitutive de tout phénomène de quantification. Toutes données reposent préalablement sur des opérations de conventions définies à partir de l'usage souhaité de la donnée.

La représentation de l'espace inscrite dans les données vise à répondre aux besoins opérationnels de l'institution qui les produit. La donnée ne vise pas à produire une représentation la plus fidèle possible de l'espace urbain : les emplacements de stationnement ou les adresses ne sont recensés de manière ni exhaustive, ni représentative⁵. Ne sont listées que celles sur lesquelles la communauté urbaine de Lyon doit agir dans le cadre de ses missions de service public. Dès lors, en fonction du découpage institutionnel et

¹ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (18 février 2015)

² Source : Entretien avec Etienne, administrateur de données, Grand Lyon (18 février 2015)

³ L'adresse postale doit en outre répondre à la norme AFNOR afin de pouvoir être traitée par les machines de tri. A noter toutefois, que La Poste dispose de plusieurs bases de données adresses selon ses métiers (tri, distribution, livraison, publicité, etc.).

⁴ Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

⁵ L'exhaustivité est la caractéristique des opérations de recensement qui visent à lister l'ensemble des entités, tandis que la représentativité est une notion statistique caractérisant la relation entre un échantillon et la population représentée (elle présuppose la connaissance d'une population de référence). Ces deux principes dépeignent deux générations de méthodes de représentation du monde. Cf. Dominique Boullier, « Vie et mort des sciences sociales avec le big data », *Socio*, 2015, n° 4, p. 19-37.

d'arrangements locaux, les données peuvent être très incomplètes à propos d'un objet particulier. La représentation inscrite dans les données est ainsi intimement liée à l'usage qui en sera fait. Difficile alors de concevoir séparément la fabrique et l'usage des données.

2) Des représentations partielles de l'espace urbain

Outre la pluralité des *visions* sur la ville, l'analyse des représentations véhiculées par les données de la Métropole de Lyon souligne leur caractère partiel. Par exemple, les informations sur le trafic routier précédemment présentées ne sont pas disponibles sur l'ensemble des voies de circulation gérées par la communauté urbaine. « *L'agglo globalement, c'est pas sur tous les axes, c'est les axes principaux*¹ » En effet, de nombreux tronçons sont peu congestionnés, inutile d'y installer des boucles inductives au coût élevé. Dès lors, l'espace sur lequel agissent les régulateurs du trafic routier diffère du territoire institutionnel de la communauté urbaine pour ne prendre en compte que les voiries structurantes.

En outre, le Grand Lyon récupère de l'information sur la circulation routière sur des voiries qui ne relèvent pas de sa domanialité : périphérique, rocade est, réseau autoroutier. « *Parce qu'on ne va pas se limiter et s'arrêter à nos frontières territoriales, parce que les déplacements, les gens se déplacent d'un point a à un point b, ils s'en foutent de savoir si c'est communal, départemental ou communautaire. Mais par contre, on ne va pas au delà, plus au nord, ça doit être le noeud de Anse, et au sud le noeud de Ternay, un peu au dessus de Vienne*². » Pour avoir cette « vision globale³ », il récupère pour cela de l'information qui lui est fourni par le centre de régulation CORALY⁴. L'obtention de données provenant de producteurs tiers permet à la Métropole de Lyon d'élargir son périmètre d'intervention au delà de ses limites institutionnelles.

Ainsi, les représentations de l'espace urbain inscrites dans les données de circulation routière sont à la fois partielles – seules certaines voies de circulation sont représentées – et

¹ Source : entretien avec Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

² Source : entretien avec Maïté, responsable du pôle Infotrafic, Grand Lyon (12 mars 2014)

³ Source : entretien avec Bruno, responsable du PC Criter, Grand Lyon (20 mars 2015)

⁴ Lancé en 1990 par la direction départementale de l'équipement du Rhône (DDE), CORALY (Coordination et Régulation du trafic sur les voies urbaines Rapides de l'Agglomération Lyonnaise) réunit les différents gestionnaires du réseau des voies rapides urbaines de l'agglomération lyonnaise afin d'optimiser la gestion du trafic (le Ministère de l'équipement et trois sociétés concessionnaires d'autoroutes : APRR, ASF, AREA). En 1999, le Grand Lyon devient partenaire de CORALY.

extensives – des voies en dehors du territoire du Grand Lyon sont incluses. La donnée repose sur des conventions, qui déterminent ce qui doit être ou non représenté. Ces conventions sont définies en fonction de l'usage que l'on souhaite faire de la donnée (ici la régulation routière).

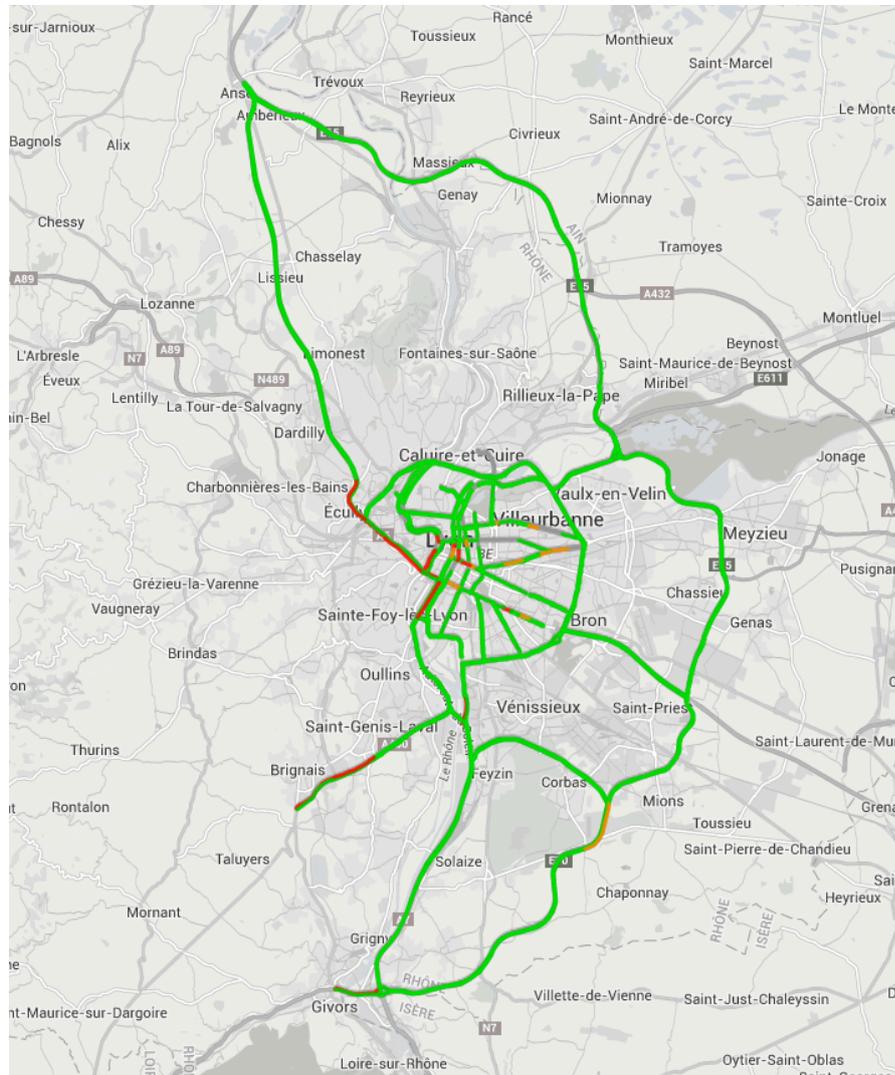


Illustration 12 : Capture d'écran du site d'information Onlymoov.com

Le périmètre sur lequel le Grand Lyon dispose d'une information sur la circulation routière est à la fois plus large que son territoire institutionnel, tout en ne représentant qu'une part minime de son réseau de voirie.

3) Les données reposent sur des conventions

La diversité des métiers et des activités engendre une variété des représentations de l'espace urbain. La pluralité des points de vue sur la ville met en lumière les conventions sur lesquelles repose toute opération de quantification. En effet, les données ne reflètent pas plus l'espace urbain que les statistiques ne mesurent les faits sociaux. Il est nécessaire de

s'éloigner d'une « épistémologie naïvement réaliste¹ » selon laquelle les objets représentés préexisteraient à leur mesure. Au contraire, la quantification nécessite, préalablement à toute mesure, une convention d'équivalence qui définit ce qu'il convient de mesurer. Les conventions d'équivalence permettent de penser dans un même cadre des cas hétérogènes. Cette opération permet le dépassement des particularités, c'est-à-dire le passage de l'individuel au collectif; ou au contraire, d'exclure certaines entités de l'espace d'équivalence. Comme l'indique Alain Desrosières : « quantifier, c'est convenir puis mesurer² ».

La construction des conventions est étroitement liée à l'usage qui sera effectué de la donnée. En fonction de leurs besoins, les producteurs construisent des modèles de données qui explicitent ce qui sera représenté et ce qui ne le sera pas. Le caractère conventionnel de la mise en données et son inscription dans une perspective d'action expliquent la variété des représentations de l'espace urbain. Plusieurs données sur un objet similaire peuvent ainsi coexister au sein d'une même institution.

Par exemple, il y a des services qui produisent une donnée de mesure de la largeur des trottoirs à des fins qui sont différentes. La propreté le fait pour des raisons de propreté, la voirie le fait pour des raisons de normes, etc. Donc, on va retrouver au Grand Lyon deux services différents qui vont faire de l'acquisition de données, chacun avec des procédés différents et des niveaux de qualité différents. Et quand on suggère que l'un utilise la donnée de l'autre : « Ah bah, non, ce n'est pas possible ! »³.

La direction de la voirie et la direction de la propreté disposent chacune d'un jeu de données sur la largeur des trottoirs. Toutefois, en fonction des spécificités de leurs métiers, ces jeux de données sont produits selon deux conventions de mesure. Ce cas est loin d'être isolé au sein de la Métropole de Lyon, où la pluralité des vues sur la ville est tout aussi diverse que les compétences de l'institution. C'est que la donnée est une médiation, au sens fort où l'entend Antoine Hennion⁴. Elle n'est pas un simple intermédiaire mettant en relation les services de la collectivité et la ville sur laquelle ils agissent. La donnée fait émerger de nouvelles entités : elle propose conjointement de nouvelles définitions des « métiers » et de l'espace urbain. La donnée crée des différences tant entre les directions de la Métropole de Lyon qu'au sein de l'objet trottoir. Si les directions de la propreté et de la voirie se refusent à

¹ A. Desrosières, *Pour une sociologie historique de la quantification*, op. cit., p. 10.

² *Ibid.*

³ Source : Journal de terrain, réunion avec Tom (20 mars 2014)

⁴ Antoine Hennion, *La passion musicale. Une sociologie de la médiation*, Paris, Métailié, 1993.

utiliser une donnée commune de largeur de trottoirs, c'est précisément parce qu'elles mènent des politiques indépendantes aux finalités distinctes pour lesquelles l'objet trottoir ne signifie pas la même chose.

Toutefois, si elle révèle la richesse et la pluralité du monde, le corolaire de cette création de différences est qu'en créant de l'incommensurabilité, elle empêche l'établissement d'une représentation commune de l'espace. Cela se reflète sur la politique publique menée comme le révèle l'exemple des transports urbains.

B. Les effets politiques des conventions : transports et mobilité urbaine

Le choix des conventions de représentation de l'espace urbain ne peut être réduite à une dimension purement technique. Toute construction de taxonomie est intrinsèquement politique¹. Les conventions d'équivalence permettent de comparer (voir ensemble) et par là, d'associer et de dissocier des éléments. Comme le souligne Alain Desrosières : « *Postuler et construire un espace d'équivalence permettant la quantification, et donc la mesure, est tout à la fois politique et technique. Il est politique en ce qu'il change le monde : comparer les roturiers et les nobles préfigure la nuit du 4 août, comparer les noirs et les blancs appelle l'abolition de l'esclavage, comparer les femmes et les hommes appelle le suffrage vraiment universel incluant les femmes* »². Les conventions d'équivalence ont des effets sociaux et cognitifs, dès lors, la quantification « *ne fournit pas seulement un reflet du monde, mais elle le transforme en le reconfigurant autrement* »³.

Le secteur des transports est un exemple intéressant des effets sociaux et cognitifs des conventions de représentation de l'espace urbain. En effet, les représentations des réseaux de transport en commun et de transport routier reposent sur deux conventions incommensurables. La représentation du réseau de transport en commun est topologique, la représentation du

¹ Sur le caractère politique de la construction d'une classe d'équivalence, on pourra se reporter aux travaux suivants : Geoffrey C. Bowker et Susan Leigh Star, *Sorting Things Out. Classification and Its Consequences*, Cambridge Mass., MIT Press, 1999 ; Laurent Thévenot et Alain Desrosières, *Les catégories socio-professionnelles.*, 5e éd., Paris, La Découverte, 2002, 121 p ; Cécile Brousse, « Définir et compter les sans-abri en Europe : enjeux et controverses », *Genèses*, 2005, vol. 1, n° 58, p. 48-71.

² A. Desrosières, *Pour une sociologie historique de la quantification*, *op. cit.*, p. 13.

³ A. Desrosières, *Prouver et gouverner*, *op. cit.*, p. 39.

réseau de transport routier est topographique¹. Si elles sont construites pour réguler ces réseaux, ces deux modalités de représentation empêchent de penser ensemble ces différents modes de transport de l'agglomération. Cet exemple des transports urbains met en avant deux agencements associant étroitement des modes de pensée, de représentations et d'actions².

Transport routier et transports en commun, c'est deux mondes différents, ils ne se parlent jamais entre eux, c'est des cultures hyper différentes. L'industrie automobile ne parle pas avec l'industrie de transport public. Et les exploitants de transport routier parlent peu avec les exploitants de transport public. C'est vraiment deux... C'est en parallèle³.

Comme le souligne cet interlocuteur, les politiques de transports urbains se caractérisent par une segmentation importante entre les modes de transport⁴ : les transports en commun d'un côté, la voiture de l'autre. Cette segmentation se reflète dans les modalités de représentation des réseaux de transports. Les données, produites pour réguler les flux afin d'éviter tout phénomène de congestion, reposent sur des conventions hétérogènes comme le démontrent les cas des réseaux routiers et de transports en commun de l'agglomération lyonnaise. Une vision monomodale a longtemps été la norme : chacun de ces réseaux est pensé, représenté et régulé de manière indépendante. Les données de circulation routière sont produites pour fluidifier les flux de circulation en pilotant les feux de circulation. Les données du réseau de transport en commun visent à assurer des flux réguliers de dessertes, qui varient selon les lignes, les événements ou la période d'exploitation.

¹ La distinction entre topologie et topographie pointe la différence entre deux métriques spatiales, l'une caractérisée par la continuité et la contiguïté (*topographique*), l'autre par la discontinuité et la connexité (*topologique*). Jacques Lévy et Michel Lussault (eds.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris, Belin, 2003, p. 927-928.

² Ce travail s'inspire des travaux de Desrosières qui met en avant cinq configurations articulant une manière de penser la société, d'agir sur celle-ci et de la représenter par la statistique.

Alain Desrosières, « Historiciser l'action publique. L'État, le marché et les statistiques. » dans Pascale Laborier et Danny Trom (eds.), *Historicités de l'action publique*, Paris, PUF, 2003, .

³ Source : Entretien avec A., Chargé de mission Mobilité urbaine au Grand Lyon, 19 janvier 2015

⁴ La segmentation du secteur est renforcée par la variété des institutions organisatrices et exploitantes. Les réseaux de transport routier sont différenciés selon les institutions en charge de leur gestion et de l'échelle de leur périmètre d'intervention : en fonction de la domanialité de la voirie, interviennent les communes, l'intercommunalité, le département, l'Etat ou encore les sociétés autoroutières. Les transports en commun sont à la charge du l'Autorité organisatrice des transports urbains - le SYTRAL dans le cas de l'agglomération lyonnaise - pour les réseaux urbains (métro, tramway, bus, funiculaire), du département pour les cars interurbains et de la région pour le réseau ferroviaire. Ces réseaux sont en outre exploités par différents opérateurs sous différentes modalités de délégation de service public. Enfin, le développement des modes doux - marche et vélo - est aujourd'hui sous la responsabilité du Grand Lyon. Cette sectorisation est encore amplifiée avec l'émergence de nouveaux services de mobilité, mis en place par des acteurs publics ou privés, caractérisés par le partage de mode de transport (covoiturage, voiture et vélo en libre-service, etc.).

Aujourd'hui, les politiques de transport urbain tendent à être envisagées sous l'angle de la mobilité urbaine. Plutôt que de penser transport en commun et transport routier l'un contre l'autre, la mobilité urbaine les envisage de manière complémentaire.

Il faut arrêter de penser en silos : il y a les transports en commun d'un côté, la voiture de l'autre, etc. Le report modal ne se fera pas de la voiture vers les TC de manière exclusive. Cela ne correspond pas aux besoins des habitants ni à la capacité d'absorption des TC¹.

Pour mettre en œuvre ce nouveau paradigme, il est nécessaire de voir émerger une nouvelle représentation des modes de transport qui intègre transport en commun et transport routier. « *Fondamentalement, il faut être capable de fournir de manière intégrée de l'information sur l'ensemble des offres de mobilité, et d'accompagner le voyageur dans son déplacement*². » Toutefois, les représentations de ces modes de transport reposent sur différentes conventions. Les données de transport en commun et les données de transport routier ne sont pas deux points de vue sur une même totalité (la mobilité) que l'on pourrait relier. L'espace du transport en commun et l'espace du transport routier sont aujourd'hui incommensurables.

Entre transports en commun et réseau routier, aujourd'hui, il n'y a pas de référentiel commun. Parce que tout simplement les normes de données ne le permettent pas. C'est-à-dire qu'un réseau routier n'est pas décrit de la même manière qu'un réseau de transport en commun. Le réseau routier, c'est des tronçons avec des positions x;y etc. Un réseau de transport en commun c'est des arrêts avec des lignes qui passent à tel arrêt, tel arrêt, tel arrêt. Une ligne passe de tel arrêt à tel arrêt. Et donc, il n'y a pas de connexion structurelle entre les deux. On ne peut pas faire de liens entre des réseaux de transports en commun et des réseaux routiers. Donc in fine, faire des liens entre une information sur la rue Paul Bert est encombrée, je sais que sur la rue Paul Bert, il y a la ligne 43, 17 et je sais pas quoi, et que donc je peux transmettre cette information aux gens du SYTRAL qui peuvent aussi l'intégrer dans leur système de gestion de trafic pour recalculer les temps de parcours. Aujourd'hui, on ne peut pas le faire. L'information existe mais ils ne savent pas l'utiliser, ils ne peuvent pas l'utiliser³.

Les différences de modélisation des réseaux routiers et de transports en commun pointées dans cet exemple sont révélatrices à double titre. D'une part, chaque acteur représente l'espace urbain en vue d'un usage déterminé. Les transports en commun de surface (bus, tramway) empruntent certes des tronçons de voirie, mais, au sein des données, ils ne

¹ Source : Interview de Jean Coldefy, chef de projet à la direction de la voirie, Grand Lyon (1^{er} novembre 2012). Disponible en ligne : <http://www.millenaire3.com/interview/services-de-mobilite-et-donnees-publiques> (Consulté le 15 décembre 2015).

² Source : Interview de Jean Coldefy, chef de projet à la direction de la voirie, Grand Lyon (1^{er} novembre 2012). Disponible en ligne : <http://www.millenaire3.com/interview/services-de-mobilite-et-donnees-publiques> (Consulté le 15 décembre 2015).

³ Source : Entretien avec Harald, Chargé de mission Mobilité urbaine, Grand Lyon (19 janvier 2015).

sont pas affectés à un tronçon puisque cette information n'est pas utile pour les gestionnaires de réseau. L'exploitation d'un réseau de transport en commun ne nécessite qu'une représentation topologique constituée du positionnement des arrêts et des lignes associées à chacun de ces arrêts. A contrario, les exploitants du réseau routier représentent topographiquement leur réseau afin de connaître le taux de congestion pour chacun des tronçons de voirie. Les normes de représentation des réseaux sont des constructions sociales qui reflètent les différences de territorialités entre transport en commun et transport routier : dans les deux cas, on régule des flux, mais on ne le fait pas de la même manière. Les données de chacun de ces exploitants sont incompatibles : ils ne parlent pas le même langage. Des référentiels communs peuvent être mis en œuvre entre les différents réseaux de transports en commun, mais pas entre le routier et le transport en commun. Ainsi, s'il existe des représentations des réseaux de transport, il n'existe pas une représentation métropolitaine de la mobilité.

*

* *

Ainsi, la donnée attache des « métiers » à la ville d'une façon particulière. La représentation de la ville ancrée dans les conventions est spécifique à un acteur et est toujours orientée vers une action. Les données de circulation sont produites par le Grand Lyon pour fluidifier le réseau routier. Les données de transport en commun sont produites par le SYTRAL pour permettre la desserte régulière des arrêts. De ces attachements particuliers émergent des agencements qui associent étroitement des manières de penser, d'agir et de représenter l'espace urbain. Si ces agencements permettent de réguler indépendamment chacun de ces réseaux de transport, ils empêchent de mener une politique de mobilité qui les penserait ensemble. Ainsi, en étant attachées étroitement à des situations et des finalités, ces données « métiers » limitent l'émergence d'une représentation commune de l'espace urbain.

C. Conclusion : les totalisations orientées de l'espace urbain

Alain Desrosières pointait l'importance de la dimension scientifique de la statistique en tant qu'outil de preuve dont les méthodes de production visaient à assurer la représentativité de l'objet représenté¹. Le travail de quantification présenté ici ne s'inscrit pas dans ce

¹ Alain Desrosières, « La statistique, outil de gouvernement et outil de preuve » dans *Pour une sociologie historique de la quantification*, Paris, Presses de l'École des Mines, 2008, p. 7-20.

paradigme. Si elles sont des dispositifs de connaissance, les données produites par les agents du Grand Lyon n'ont aucune prétention à la représentativité de l'objet représenté. Elles sont davantage des outils destinés à l'action que des outils de preuve ; elles se distinguent en cela des statistiques publiques traditionnellement étudiées par la sociologie de la quantification.

Impossible alors de dissocier la donnée de son usage. Ces représentations de l'espace urbain sont tournées vers une action particulière : ramasser les déchets, réguler la circulation routière, assurer le service de transport en commun, etc. Toute donnée est un instrument d'action publique, au sens où l'entendent Pierre Lascoumes et Patrick Le Galès, « *un dispositif technique à vocation générique porteur d'une conception concrète du rapport politique/société et soutenu par une conception de la régulation* »¹. Cette définition souligne le caractère intrinsèquement politique des instruments. S'intéresser au processus de quantification ne signifie donc pas se cantonner à évoquer ses aspects techniques, mais à s'attacher à décrire les représentations sociales et les relations de pouvoir qui s'inscrivent dans ces modalités « techniques ».

Ce constat nous éloigne d'une vision positiviste des données envisagées comme des outils objectifs et neutres de représentation d'une réalité qui leur préexisteraient. Si tel était le cas, mettre en donnée un espace reviendrait à *révéler* cette réalité selon un processus mécanique et objectif qui ne peut modifier le contenu de ce qui est révélé. Les données reposent au contraire sur des conventions qui définissent ce qu'il convient de mettre en donnée. Comme le rappelle Alain Desrosières, « *L'information statistique ne tombe pas du ciel comme pur reflet d'une "réalité" antérieure à elle. Bien au contraire elle peut être vue comme le couronnement provisoire et fragile d'une série de conventions d'équivalence entre des êtres qu'une multitude de forces désordonnées cherche continuellement à différencier et à disjoindre*² ». Ce travail de construction de conventions est étroitement lié aux usages que l'on souhaite faire des données. Cette posture conventionnaliste nous éloigne de toute neutralité des données : elles incorporent une représentation limitée du monde³. Les raisons

¹ Pierre LASCOUMES, Patrick LE GALÈS. Introduction : L'action publique saisie par ses instruments. In *Gouverner par les instruments*. Paris : Presses de Sciences Po, 2005.

² Alain Desrosières, « Discuter l'indiscutable. Raison statistique et espace public. », *Raisons Pratiques*, 1992, n° 3, p. 131-154.

³ Le caractère politique des conventions se manifeste par les controverses qui apparaissent lors des opérations de construction des classes d'équivalence, où chaque groupe d'acteurs tente d'imposer la définition de l'objet représenté qui se rapproche le plus de ses intérêts. Voir notamment à ce sujet : G.C. Bowker et S.L. Star, *Sorting Things Out. Classification and Its Consequences*, *op. cit.* ; C. Brousse, « Définir et compter les sans-abri en Antoine COURMONT – « Politiques des données urbaines » - Thèse IEP de Paris – 2016

qui président en cela sont en premier lieu pragmatiques : il est impossible de tout représenter. Cette remarque peut paraître triviale, elle est pourtant fondamentale à la compréhension. La donnée n'est qu'une représentation de la ville, elle est nécessairement partielle et réductionniste. La donnée et la ville se rencontrent sans se confondre.

En définissant des espaces d'équivalence entre des entités hétérogènes, ces conventions produisent des totalisations de l'espace urbain. Toutefois, ces totalisations sont toujours partielles : les données ne représentent pas l'ensemble de l'agglomération, mais elles limitent la focale pour restreindre le *monde* aux éléments pertinents pour les actions plurielles des pouvoirs publics. L'agglomération des régulateurs du trafic routier n'est pas celle des collecteurs de déchets ou celle des chauffeurs de bus. Cela n'intéresse pas les régulateurs de connaître le trafic sur la petite voie de desserte où circulent 50 véhicules par jour, mais où l'éboueur doit aller collecter deux fois par semaine les déchets ménagers. A contrario, l'éboueur n'a que faire du périphérique, lieu central de la régulation du trafic, mais où aucun déchet n'est à collecter. Les données sont ainsi autant d'oligooptiques, des « *étroites fenêtres qui permettent de se relier, par un certain nombre de conduits étroits, à quelques aspects seulement des êtres (humains et non-humains) dont l'ensemble compose la ville* »¹. Elles ne représentent bien que peu de choses² et ne donnent à voir qu'une vision partielle du territoire de l'institution Grand Lyon. Chaque donnée représente un espace d'action particulier sur lequel interviennent seulement certains services de la communauté urbaine. Les données sont ainsi au cœur d'agencements fonctionnels qui attachent un « métier » et un point de vue sur la ville dans une visée opérationnelle. Le gouvernement de la ville est ainsi constitué d'une « *mosaïque de petits mondes qui se touchent sans s'interpénétrer* » pour reprendre les mots de Robert Park³. La spécialisation fonctionnelle de ces mondes se manifeste par leur homogénéité interne, un point de vue partagé sur la ville, et par l'hétérogénéité des représentations de l'espace urbain entre les métiers. Cette métaphore de la mosaïque est utilisée par les sociologues de l'École de Chicago pour pointer les phénomènes de

Europe », art cit ; Delphine Nivière, « Négocier une statistique européenne : le cas de la pauvreté », *Genèses*, 2005, vol. 1, n° 58, p. 28-47.

¹ Bruno Latour, « Paris, ville invisible : le plasma » dans Christine Macel, Daniel Birnbaum et Valérie Guillaume (eds.), *Airs de Paris, 30 ans du Centre Pompidou*, Paris, ADGP, 2007, p. 260-263.

² Bruno Latour et Hermant, *Paris, ville invisible, Les empêcheurs de penser en rond.*, Paris, La Découverte, 1998, 159 p ; B. Latour, *Reassembling the Social. An introduction to Actor-Network Theory, op. cit.*, p. 181.

³ Robert Ezra Park, « La ville. Propositions de recherche sur le comportement humain en milieu urbain » dans Isaac Joseph et Yves Grafmeyer (eds.), *L'école de Chicago : Naissance de l'écologie urbaine*, Paris, Flammarion, 2005, p. 125.

ségrégations spatiales et sociales. Appliquée au gouvernement de la ville, elle met en évidence le cloisonnement des différents métiers d'une même institution. Au sein de la communauté urbaine de Lyon, coexiste une pluralité de représentations de la ville qui correspondent à chacune de ses missions. Comment alors assurer un monde commun face à la multiplicité des représentations de l'espace ? Comment les acteurs peuvent-ils se coordonner s'ils ne partagent pas le même langage, s'ils ne voient pas la même ville au travers de leurs données ?

III. Les données comme outil de coordination : une représentation partagée du territoire métropolitain

Au début des années 80, la communauté urbaine de Lyon est confrontée à un souci de coordination entre ses différentes directions métiers. Chacune produit des plans et/ou des données géographiques, mais ces informations ne sont pas échangées entre directions.

On refabriquait à des tas d'endroits des documents qui existaient ou dont les données avaient été fabriquées ailleurs. [...] Il y avait un service des plans, qui s'occupait des travaux avec les géomètres, qui passait les marchés avec les géomètres, mais, dans tous les services de la Courly, chacun faisait ce qu'il voulait avec qui il voulait. Le service voirie commandait des géomètres, le service des eaux, des géomètres, chacun faisait son petit marché, personne ne s'occupait de quelqu'un d'autre. [...] Il y avait des plans de géomètres à la voirie, qui auraient pu servir à l'assainissement, mais, ils ne le savaient pas¹.

Une enquête effectuée en 1983 auprès des services techniques sur les besoins en représentation cartographique révèle l'ampleur de ces activités de production de plans et de leur « nécessité vitale² » pour l'accomplissement de leurs missions. Plus de 200 000 plans sont répertoriés au sein des services techniques, dont 25 000 sont produits par les services communautaires³. Une centaine de personnes travaille alors à leur production ou au recueil de données. L'hétérogénéité de ces plans et le faible échange d'informations entre les services

¹ Source : entretien de Jean Barboyon, entré à la Communauté urbaine en 1969, il est ingénieur en chef du centre de données urbaines du Département Développement de 1984 à 1990. Entretien disponible sur le site <http://www.40ans.grandlyon.com/?p=6415> (Consulté le 3 juin 2015).

² Source : L'évaluation du système urbain de référence de la communauté urbaine de Lyon – Rapport d'étape – ENTPE - 1989

³ Source : Rapport « L'évaluation du système urbain de référence de la Communauté urbaine de Lyon » rédigé par l'Ecole Nationale de Travaux Publics de l'Etat (ENTPE) en juillet 1989.

techniques sont la source de dépenses onéreuses d'acquisition de données, de doubles emplois dans différents services et de surcoûts des travaux dus à la mauvaise qualité des plans.

Cette faible coordination est d'autant plus problématique que les lois de décentralisation de 1982 confèrent aux structures intercommunales de nouvelles compétences en matière d'urbanisme. La communauté urbaine ne doit alors plus seulement assurer la gestion quotidienne des réseaux techniques de l'agglomération mais également définir et mettre en œuvre les politiques d'aménagement. Ces nouvelles fonctions nécessitent la coordination d'acteurs hétérogènes autour d'une représentation partagée de l'espace à aménager. Il a alors été décidé de mettre en œuvre un « système urbain de référence » (SUR): une « banque de données » fournissant une information géoréférencée commune à l'ensemble des services de la communauté urbaine de Lyon.

Le Système urbain de référence, c'est faire sur le territoire de la Courly, une banque de données, servant à tout le monde, c'est-à-dire ayant la même référence pour que tout le monde se repère aux mêmes données géographiques. Pour que le service des eaux, s'il regarde le plan, il sait que là il y a un assainissement, là il y a un groupe scolaire, là, il y a des lignes EDF, etc. Chacun avait ses données pas forcément par rapport au cadastre, pas par rapport à l'angle de l'immeuble, jamais par rapport avec la même chose. Donc, quand vous compariez, c'était impossible, d'où l'idée de remettre tout ça et de le rendre cohérent. Et d'en faire un système de référence pour tout le monde¹.

Le SUR prend la forme d'un système d'information géographique (SIG)², pionnier à l'époque pour les collectivités territoriales³, qui repose sur une architecture innovante de superposition des informations. Aux couches communes de données de référence, les services ajoutent les informations relatives à leur domaine d'activité.

¹ Source : entretien de Jean Barboyon, entré à la Communauté urbaine en 1969, il est ingénieur en chef du centre de données urbaines du Département Développement de 1984 à 1990. Entretien disponible sur le site <http://www.40ans.grandlyon.com/?p=6415> (Consulté le 3 juin 2015).

² Les systèmes d'information géographique peuvent être définis comme « *A system of hardware, software, data, people, organizations and institutional arrangements for collecting, storing, analyzing, and disseminating about areas of the earth.* » (Nicholas Chrisman, *Exploring geographic information systems*, 2nd éd., s.l., Wiley, 2002, p. 12)

Selon Joliveau, les SIG comprennent quatre éléments : une composante technologique (logiciel et matériel), une composante informationnelle (une base de données spatialisées), une composante organisationnelle et une composante méthodologique. Thierry Joliveau, « Gérer l'environnement avec des SIG. Mais qu'est-ce qu'un SIG ? », *Revue de géographie de Lyon*, 1996, vol. 71, n° 2, p. 101-110.

³ Pour un historique de l'usage des SIG au sein des collectivités locales, voir : Stéphane Roche, *Les enjeux sociaux des systèmes d'information géographique*, Paris, L'Harmattan, 2000, 158 p ; Grégoire Feyt, « Les métiers du territoire face aux technologies de l'information géographique : Babel et esperanto » dans Stéphane Roche et Claude Carron (eds.), *Aspects organisationnels des SIG*, Paris, Hermès, 2004, p. 55-70.

Le système urbain de référence c'était surtout de la géographie informatisée, on voulait cartographier, on voulait superposer des couches d'information, de façon à avoir en un seul endroit toutes les couches d'information qui correspondaient à cet endroit là. Je clique sur une rue, je vois apparaître le cadastre, les données sur la largeur de la voie, j'ai les données sur le réseau d'assainissement, qu'il y a dessous, j'ai les données sur les arbres d'alignement qui sont plantés, avec les caractéristiques de ces arbres, ils sont de telle essence, ils ont été plantés à telle époque, ils sont en bon ou mauvais état de santé, j'ai des informations sur le réseau d'eau, j'ai des informations sur la population, on avait même travaillé à l'époque avec la Cité administrative de l'Etat qui gérait les recensements de population, on avait des informations sur le PLU, le Plan d'occupation des sols. L'objectif du centre de données urbaines c'était vraiment ça, cliquer et rapidement avoir toutes les informations superposées¹.

Ces données de référence permettent ainsi l'attachement d'acteurs multiples (services du Grand Lyon, services de l'Etat, etc.) autour d'une représentation partagée de l'espace urbain. Ce cas va à l'encontre de l'argumentation mise en avant dans la partie précédente de ce chapitre qui mettait en avant les représentations partielles de l'espace urbain définies pour répondre aux besoins spécifiques d'une direction métiers. Qu'est-ce qui distingue ces données métiers de ces données de référence ?

La réponse est à rechercher dans l'ampleur de « l'investissement de forme » réalisé pour établir la convention sur laquelle repose la mise en données du monde. Plus l'investissement est conséquent, plus la donnée parviendra à attacher des acteurs aux intérêts hétérogènes autour d'une représentation commune. Les caractéristiques particulières des données de référence permettent l'émergence d'un agencement territorial associant une institution – la communauté urbaine de Lyon – à un territoire.

A. Données de référence et données métiers : un investissement de forme d'ampleur différente

Les systèmes d'information géographique sont constitués de couches d'informations superposées. Cette superposition n'est possible que par le partage d'un référentiel géographique à partir duquel chaque producteur produit ses données spécifiques. Le SIG est ainsi composé de deux types de données : les données métiers et les données de référence comme l'illustre ce schéma de présentation du SIG de la communauté urbaine de Lyon.

¹ Entretien de Catherine David, entré à la communauté urbaine en 1987, ingénieur au sein du service de l'assainissement, en charge de la définition et de la mise en place du système urbain de référence jusqu'en 1992. Entretien disponible sur le site <http://www.40ans.grandlyon.com/?p=6415> (Consulté le 3 juin 2015).

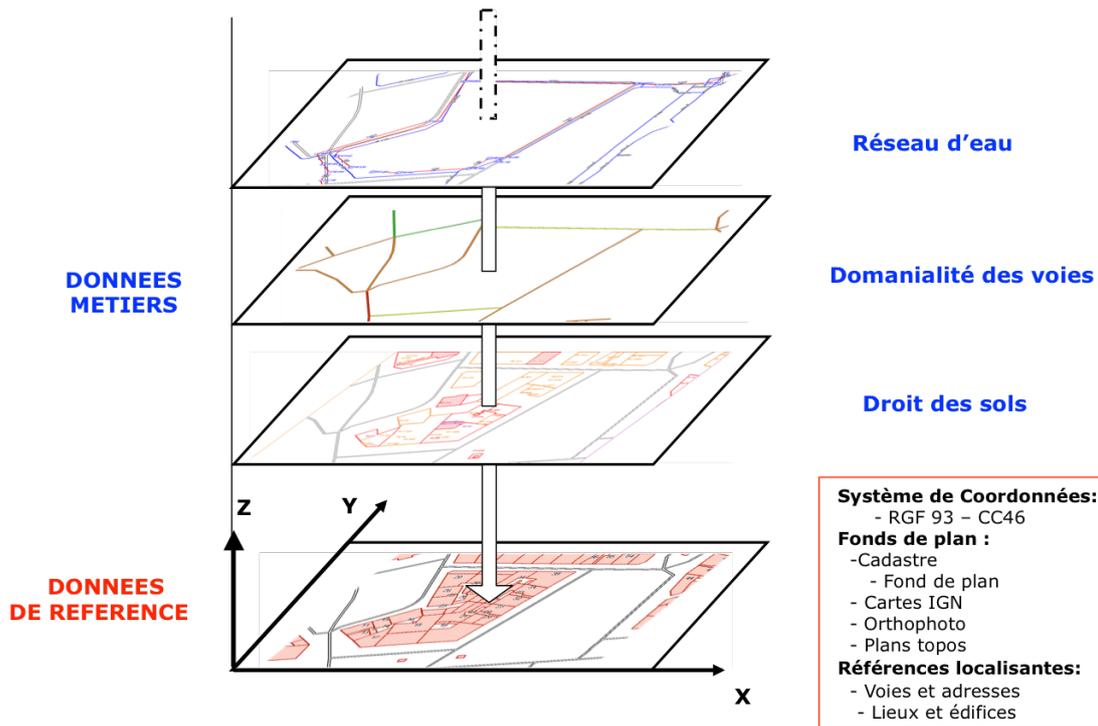


Illustration 13 : Les « couches » du SIG articulées par un référentiel unique de localisation¹

Les données de référence (ou *référentiels géographiques*) sont des éléments cruciaux de tout système d'information géographique puisqu'ils permettent la mise en relation des différentes couches de données. Définies comme l'ensemble des « *éléments susceptibles de contribuer à la contextualisation d'informations thématiques relevant de problématiques, de périmètres ou de métiers particuliers* »², les données de référence établissent une logique commune entre les différents moyens de localisation. Revenir sur les modalités de production de ces données de référence permettra d'explicitier ce qui les distingue des données métiers. Si on oppose régulièrement données métiers et données de référence en fonction de leur articulation à un usage particulier, il semble plus juste de souligner que la distinction entre ces données provient de l'ampleur de l'investissement de forme sur lequel reposent leurs conventions.

¹ Source : Document interne de présentation du SIG du Grand Lyon – Mis à jour en octobre 2010

² Grégoire Feyt et Sylvie Lardon, « L'émergence de nouveaux référentiels pour dire le territoire : vers une complémentarité des points de vue ou un "combat des cartes" ? », Montpellier, 2010.

1) Données métiers et données de référence : une distinction par les usages ?

Les données de référence et les données métiers correspondent à deux modes de production de la donnée géographique¹. Les premières sont constituées d'informations à caractère général relatives à la description du territoire. Les secondes sont des données thématiques correspondant aux domaines d'activités de chacune des directions du Grand Lyon. L'articulation à leurs usages distingue les données de référence des données métiers. Les premières s'inscrivent dans une logique de description du territoire, tandis que les secondes répondent à un besoin d'action sur le territoire. Le parallèle avec les deux sources de la statistique publique identifiées par Alain Desrosières est ici intéressant². Les données de référence correspondraient aux « enquêtes », telles que le recensement de la population, dont l'objectif est d'explorer l'état de la société avec un souci d'exhaustivité et une recherche de scientificité³ dans la représentation du monde. La représentation du territoire doit être la plus fidèle possible sans préjuger des usages qui en seront faits. Les données de référence sont ainsi susceptibles de s'autonomiser de leurs usages comme l'indique leur responsable.

Finale­ment, mon fond de plan est constitué de plein de couches, dont les utilisations vont aller d'une échelle très grande à une échelle très petite. [...] L'utilisateur peut piocher là dedans les couches qui l'intéressent et se fabriquer, un peu comme une pizza, prendre ce qu'il veut comme garniture, et prendre son fond de plan qui va lui permettre de répondre à un usage donné, que nous on n'a pas forcément identifié. **On ne préjuge pas de l'usage qui peut être fait de ces informations là**⁴.

A contrario, les données métiers peuvent être comparées aux « sources administratives » qui sont inscrites dans l'action quotidienne de l'Etat, et en particulier dans ses activités de gestion. Elles répondent à des usages bien identifiés pour lesquelles elles sont spécifiquement produites, sans souci de représentativité. Ces données visent plus à permettre et garantir l'action qu'à refléter des « réalités ». Elles sont étroitement liées à des usages, on choisit de ne représenter que ce qui est nécessaire à l'action.

¹La distinction entre données métiers et données de référence n'est pas spécifique au Grand Lyon, mais elle est courante au sein de la géomatique. Jean Denègre et François Salgé, *Les systèmes d'information géographique*, Que sais-je ?, Paris, Presses Universitaires de France - PUF, 2004, 128 p.

²A. Desrosières, « Décrire l'Etat ou explorer la société : les deux sources de la statistique publique », art cit.

³ La recherche de scientificité repose sur la démarche méthodologique mise en œuvre qui vise à objectiver ces représentations sociales.

⁴ Source : Entretien avec Etienne, administrateur des données de référence de la Communauté urbaine de Lyon (26 novembre 2013).

La distinction entre ces deux familles de données géographiques repose sur leur articulation aux usages, les données de référence semblant être décorrélées d'un usage particulier¹. En visant la représentativité du territoire, ces données visent à atteindre un degré de généralité permettant de répondre aux besoins de tous les services. Cela se reflète dans le souci d'exhaustivité et de neutralité qui est poursuivi : il s'agit de ne pas privilégier un usage ou un territoire au détriment d'un autre, mais de fournir une représentation homogène, exhaustive et la plus précise possible du territoire de la Métropole de Lyon.

En tant que producteurs de données fédérales, on se doit d'avoir un socle exhaustif et complet sur toutes les communes, de même niveau de qualité. [...] Notre doctrine, c'est ça : on met à disposition des données sur l'ensemble du Grand Lyon de même qualité. La commune de Chassieu aura le même niveau de qualité que la commune de Lyon ou de Poleymieux².

Les photographies aériennes, qui constituent une source essentielle pour la production de toutes les données de référence, illustrent cette recherche de qualité. Chaque campagne est l'objet d'un cahier des charges rigoureux qui décrit précisément les conditions de prise de vue afin d'obtenir une représentation du territoire de plus en plus précise.

A chaque campagne de prise de vue aérienne, on est plutôt allé en améliorant la qualité. Soit en terme de résolution, soit de précision. On a toujours augmenté. La première campagne d'orthophoto date de 93, on en a fait depuis en 97, 2003, 2007, 2009 et 2012. [...] On est depuis 97 sur une résolution de 16 centimètres. C'était à l'époque révolutionnaire, personne n'avait fait sur une telle superficie une ortho avec un pixel de 16cm. On est passé à 10 cm en 2012. [...] Et on a gagné à chaque fois en précision. Sur la dernière ortho, on estime et on a mesuré que la précision, elle est autour de deux pixels, donc on est à 20 centimètres. Et ça, c'est super bien. Sur Google Earth par exemple, ils peuvent avoir la même résolution que nous, par contre, ça peut être décalé de 50 centimètres, voire de plusieurs mètres à certains endroits. Ils n'ont pas cette recherche de qualité³.

En répondant à une définition techniciste de la qualité, qui recherche plus de précision par une amélioration continue des outils de collecte, les données de référence s'inscrivent dans une épistémologie réaliste inspirée des sciences de la nature⁴. Cette précision ne peut en effet être jugée qu'à l'aune d'une réalité préexistante à la mesure. En représentant un territoire

¹ Cette distinction est renforcée par la division organisationnelle entre les producteurs des données de référence et des données métiers. Les premiers sont rattachés à une direction transversale, la délégation générale aux ressources (DGR), tandis que les seconds appartiennent aux directions sectorielles (voirie, propreté, eau, etc.).

² Source : Entretien avec Romain, administrateur de données, Grand Lyon (21 novembre 2013)

³ Source : Entretien avec Etienne, administrateur des données de référence de la Communauté urbaine de Lyon (26 novembre 2013).

⁴ Le « réalisme » ou « naturalisme » est « la doctrine qui soutient que les phénomènes existent avant même que d'être catégorisés et indépendamment de tout travail de catégorisation. » Cyril Lemieux, « Peut-on ne pas être constructiviste ? », *Politix*, 2012, vol. 4, n° 100, p. 172.

qui existe antérieurement à la mesure, les données de référence apparaissent en position d'extériorité par rapport aux usages sociaux. Cette extériorité leur confère une visée universaliste, ce qui leur permet d'être un langage commun reconnu et accepté par l'ensemble des acteurs de l'agglomération.

*

* *

Cette autonomisation des données de référence à un usage particulier se manifeste dans le rattachement du service produisant ces données à une direction ressource à vocation transversale. En 1997, le Centre des données urbaines, service à l'origine et en charge du SUR, devient le service des systèmes d'information communautaires. Il quitte le Département du développement urbain pour être rattaché à la Direction de la logistique et des bâtiments. Quelques années plus tard, il change à nouveau de direction pour être placé au sein de la direction des systèmes d'information et des télécommunications, nouvellement créée. Ces changements organisationnels manifestent l'autonomisation du service SIG de la direction de l'urbanisme pour s'adresser à l'ensemble des directions métiers de la communauté urbaine. Depuis la constitution des premières bases de données « métiers », le SUR n'est en effet plus uniquement un outil au service des politiques d'urbanisme et de développement urbain, mais s'adresse à l'ensemble des acteurs intervenant sur le territoire de la communauté urbaine.

Les « données métiers » sont produites au sein de chaque direction opérationnelle du Grand Lyon. A partir des années 90 et avec le développement de la micro-informatique¹ et des logiciels bureautiques de système d'information géographique (SIG), elles constituent sur la base de ces données de référence leurs propres données géographiques dont ils ont l'entière responsabilité. Cette décennie 90 marque véritablement l'usage opérationnel des SIG. Après avoir assuré la numérisation des plans et des cartes et la gestion de ces données géographiques, l'exploitation de ces données, les analyses et les productions de cartes qui en découlent, prennent toute leur ampleur avec le développement des applications métiers et la facilité d'accès sur les postes informatiques individuels des agents de l'administration.

Au sein des directions du Grand Lyon, des applications spécifiques de saisie et de gestion de données sont créées progressivement. Des services organisationnels en charge de la

¹ En 1993, la communauté urbaine de Lyon dispose de 430 PC, 230 Macintosh et 220 terminaux passifs. En quatre ans, le nombre d'ordinateurs va être multiplié par 2,5. (Source : Cédric Polère)

production, de la gestion et de la mise à jour de ces données sont mis en place au sein des directions. Ces services « métiers » ont une autonomie totale du centre de données urbaines.

C'est ainsi que les services d'urbanisme ont pu, tout à la fois, numériser les six plans d'occupation des sols, les mille servitudes qui couvrent le territoire, les alignements des voies nouvelles. Ils ont pu également positionner chaque demande de permis de construire, de permis de démolir, de certificat d'urbanisme, de lotissement, suivre l'évolution du marché de l'immobilier de manière très fine, traiter géographiquement les données socio-économiques du recensement général de la population. La direction de l'Action foncière a dressé de manière détaillée l'atlas des quatre mille propriétés communautaires. Les Services urbains ont quant à eux numérisé les 2500 kilomètres du réseau d'égout et peuvent ainsi mieux connaître la qualité de celui-ci, tronçon par tronçon, et en améliorer l'entretien. Le réseau d'eau est également en cours de numérisation. Ils ont aussi positionné plus de 37 000 points lumineux de l'éclairage de la ville de Lyon ainsi que les feux tricolores. De même, 52 000 arbres d'alignement avec leur essence et leur taille, les accidents constatés et les travaux programmés sur la voirie sont localisés¹.

Les premiers services à s'approprier cet outil et à y ajouter leurs propres données sont l'urbanisme et l'assainissement. Les directions de l'action foncière et de la voirie vont également s'en emparer rapidement. Aujourd'hui, les principales directions opérationnelles de la communauté urbaine disposent d'un SIG. Des personnes dédiées à la production, la gestion et les mises à jour de ces données sont en place au sein des principales directions : les administrateurs métiers. Leurs activités sont diverses : administration des données dans le SIG, saisie, mise à jour, contrôle et intégration de données, acquisition de données, développement d'applications SIG, production de cartes et de tableaux, diffusion de données, formation et assistance aux utilisateurs. Face à la diversité de ces activités, ces services géomatiques se sont développés et regroupent des agents aux profils et aux grades divers.

¹ Pierre Pijourlet, « Le système urbain de références de la communauté urbaine de Lyon » dans *Forma Urbis - Les plans généraux de Lyon : XVIe-XXe siècle*, Lyon, Archives municipales de Lyon, 1997, .

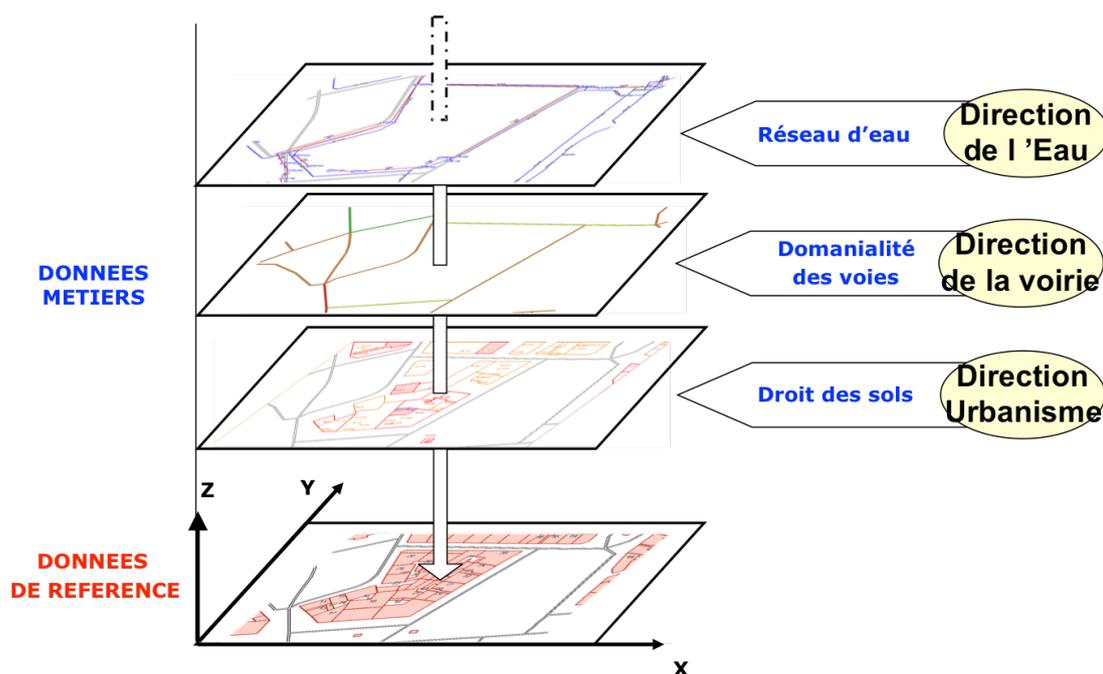


Illustration 14 : Schéma de présentation du SIG et direction métiers correspondante¹

Ce schéma de présentation du SIG de la communauté urbaine de Lyon rend manifeste l'autonomisation des données de référence de tout usage opérationnel. Alors que les exemples de données métiers sont attachés à une direction productrice, les données de références sont autonomisées : le service en charge de leur production n'est pas mentionné. Cette absence de référence à un producteur identifié contribue au « réalisme » de ces données.

*

**

Cette distinction courante entre données de référence et données métiers par le rapport à l'usage n'est toutefois pas entièrement satisfaisante. D'une part, les données métiers peuvent devenir des données de référence. On peut songer par exemple aux registres d'état civil qui sont devenus des référentiels objectivés autonomisés de leurs usages initiaux. D'autre part, lorsque l'on interroge les producteurs des données de référence sur leur travail quotidien, on s'aperçoit que si les données de référence s'inscrivent dans une logique de représentation exhaustive du territoire, elles ne sont toutefois pas totalement indépendantes des usages qui en sont fait. Les données de référence ne s'inscrivent en effet pas uniquement dans un cadre de représentation, mais également dans un cadre d'action. Des choix de représentation sont donc

¹ Source : Document interne de présentation du SIG du Grand Lyon – Mis à jour en octobre 2010

effectués en fonction des usages connus du référentiel géographique afin que celui-ci y réponde. Les données sont modélisées pour la plus grande variété d'utilisations possibles, mais non pour toutes les utilisations.

Par exemple, la donnée « Lieux et édifices » ne contient pas d'informations sur les équipements communaux (crèches, centres communaux d'action sociale, etc.) ou sur les grands équipements culturels et sportifs (opéra, musées, stades). De fait, les compétences culturelles, sportives ou sociales ne sont pas des prérogatives de la communauté urbaine, les services n'en n'ont a priori pas usage, ces champs ne sont donc pas renseignés. Les données intègrent un cadre d'action dont les limites se font sentir lorsque les compétences de l'institution évoluent.

On a saisi ces thèmes en fonction des besoins des services internes essentiellement. Alors on se pose des questions sur ce qui va arriver avec la métropole, et les compétences qu'on ne gère pas du tout au Grand Lyon et qu'on va avoir avec la Métropole¹.

A l'instar des données métiers, les données de référence reposent sur des conventions définies par rapport à des usages (perçus, potentiels ou réels). La qualité de la donnée ne repose alors plus sur des critères techniques assurant la précision et la fiabilité de l'objet représenté, mais sur une conception sociale relevant de la pertinence de la donnée, soit sa correspondance aux besoins des utilisateurs². Par exemple, l'élargissement des compétences de l'institution conduit ce producteur à s'interroger sur les éléments qu'il devra ajouter dans sa base afin de couvrir les besoins des services de la nouvelle institution. De manière similaire, il s'interroge sur la pertinence à maintenir une représentation homogène du territoire alors que les usages ne sont pas identiques.

J'essaye de couvrir tous les territoires de la même façon, alors que finalement les enjeux sont pas les mêmes et qu'il y a forcément une différenciation à faire sur les zones urbanisées et des zones rurales pour lesquelles le besoin de précision et de qualité de données est peut être moindre, même si le besoin de couverture existe. Jusqu'à présent j'ai toujours dit « il faut que nos bases de données de référence, elles soient partout pareilles, il n'y a pas de raison de favoriser telle ou telle chose ». Dans les faits, ce n'est pas si vrai que ça, il faut se poser les bonnes questions. Il y a des zones à enjeux, c'est comme ça, il faut faire avec. Des grands aménagements type Carré de soie ou

¹ Source : Entretien avec Etienne, administrateur des données de référence de la Communauté urbaine de Lyon (26 novembre 2013). Entretien réalisé avant le 1^{er} janvier 2015, date à laquelle la Communauté urbaine de Lyon et le Département du Rhône ont fusionné leurs compétences sur le territoire du Grand Lyon pour former la Métropole de Lyon.

² Sur les « qualités » d'un instrument quantitatif et la distinction précision/pertinence, voir Alain Desrosières, « Les qualités des quantités », *Courrier des statistiques*, 2003, n° 105-106, p. 51-63.

Confluence¹, tu ne peux pas exactement les traiter de la même manière que le dessus de Curis-au-Mont-d'or² où il ne se passe rien³.

La production des données de référence ne s'effectue pas de manière autonome aux usages. Tout comme les données métiers, elles reposent sur des conventions qui définissent ce qui doit être représenté. Elles s'inscrivent alors dans une épistémologie constructiviste⁴, puisqu'elles ne mesurent pas une réalité préexistante, mais elles la font exister. Dans cette conception, le territoire ne préexiste pas à sa représentation. Il doit en outre être envisagé de manière dynamique, puisque, comme le souligne l'évolution des données de référence, le territoire de la communauté urbaine de Lyon n'est pas identique au territoire de la Métropole du Grand Lyon.

Contrairement à l'idée commune d'une autonomisation des données de référence d'usages prescrits, celles-ci sont également produites pour répondre à certains besoins. « *Tout à la fois, elles reflètent la réalité et elles instituent celle-ci* »⁵. Comment expliquer ces discours contradictoires sur une même donnée ?

Ce paradoxe apparent met en avant la fausse opposition entre réalisme et constructivisme. Si l'on suit les discours des producteurs, les données sont tout à la fois réalistes et constructivistes. La question n'est alors pas tant d'opposer ces deux épistémologies que d'étudier dans quelles circonstances elles sont tour à tour mobilisées par les acteurs. Comme le souligne Alain Desrosières, « *Le choix entre les deux postures, réaliste ou conventionnaliste, n'est pas un choix existentiel engageant les personnes de façon permanente. Il est lié aux contraintes des situations auxquelles sont confrontés ces acteurs, selon les moments, et selon les types d'interaction avec des objets ou avec des personnes qu'ils rencontrent* »⁶. Cette réflexion nous invite à étudier plus précisément le travail des

¹ Ces deux quartiers sont l'objet de projets de réaménagement urbain de grande envergure.

² Commune de la communauté urbaine de Lyon représentant environ 0,58% du territoire et 0,08% de la population du Grand Lyon.

³ Source : Entretien avec Etienne, administrateur des données de référence de la Communauté urbaine de Lyon (26 novembre 2013).

⁴ Le constructivisme est entendu ici comme « *la doctrine selon laquelle les phénomènes descriptibles dans le monde, qu'ils soient réputés ordinairement sociaux ou naturels, n'existent pas antérieurement et extérieurement au travail accompli pour les catégoriser* ». C. Lemieux, « Peut-on ne pas être constructiviste ? », art cit, p. 172.

⁵ Souligné par Alain Desrosières. A. Desrosières, *Prouver et gouverner*, op. cit., p. 147.

⁶ A. Desrosières, « Les qualités des quantités », art cit, p. 60.

producteurs de données de référence. On s'aperçoit alors que la distinction entre données métiers et données de référence repose davantage sur l'ampleur de l'investissement réalisé pour assurer la stabilité de la représentation de l'espace qu'elles véhiculent.

2) L'ampleur de « l'investissement de forme »

Ce va-et-vient permanent entre les épistémologies réalistes et constructivistes révèle l'ampleur de « l'investissement de forme¹ » sur lequel repose les données de référence. Ces deux points de vue sur la donnée s'expriment à différentes étapes de la chaîne de la donnée. Le producteur a conscience de la « vulnérabilité » de sa donnée face aux usages auxquels elle ne peut répondre. Pour assurer la solidité des données face aux « débordements » dont elles sont l'objet², il doit adapter les conventions initiales aux évolutions des usages. Pour cela, des dispositifs de remontées d'informations auprès des utilisateurs sont mis en place.

J'essaie de maintenir à jour un tableau de suivi pour savoir ce qu'ils en font. Ça permet un peu de voir concrètement les usages, de voir comment on pourrait améliorer la donnée³. On est prêt à faire évoluer notre structure en fonction des demandes. On est là pour ça [faire évoluer les données]⁴.

Ces procédures d'ajustement et de révision de la donnée pour la faire correspondre à l'évolution des situations soulignent la « fragilité » des objets techniques que constituent les données et le soin quotidien qui leur est apporté pour garantir leur solidité⁵.

A contrario, l'utilisateur de la donnée doit pouvoir s'appuyer sur une donnée stabilisée. Le producteur procède donc à un travail de clôture, mettant en « boîte noire » les conditions techniques de production, pour donner à la donnée une apparence de stabilité et d'autonomie. La donnée répondant aux usages, les modalités de production sont invisibilisées. Hormis, le producteur et le gestionnaire de la donnée, les données sont intégrées dans une routine

¹ L. Thévenot, « Les investissements de forme », art cit.

² Les débordements correspondent aux usages imprévus qui débordent du cadre initial envisagé. Les débordements ne peuvent en effet s'appréhender qu'en fonction des cadrages qui ont été opérés préalablement : l'identification précise des acteurs concernés par la donnée et des usages qui en seront effectués. Cette notion de « cadrage-débordement » a été développée par Michel Callon pour analyser les transformations des agencements marchands. Michel Callon, « La sociologie peut-elle enrichir l'analyse économique des externalités ? Essai sur la notion de cadrage-débordement » dans Dominique Foray et Jacques Mairesse (eds.), *Innovations et performances. Approches interdisciplinaires*, Paris, Editions de l'EHESS, 1999, p. 399-431.

³ Source : Entretien avec Romain, administrateur de données, Grand Lyon (21 novembre 2013)

⁴ Source : Entretien avec Romain, administrateur de données, Grand Lyon (21 novembre 2013)

⁵ Jérôme Denis, « L'informatique et sa sécurité. Le souci de la fragilité technique. », *Réseaux*, 2012, vol. 30, n° 171, p. 161-187.

d'action, et les conventions qui ont présidé à leur production sont encapsulées dans des « boîtes noires ». Pour les utilisateurs de la donnée, les conventions sont naturalisées, la donnée est incontestable, son objet est réifié. La donnée tend à devenir pour les utilisateurs l'objet mesuré.

Les données de référence sont des objets qui sont à la fois « *construits et réels, conventionnels et solides* »¹. Ces deux points de vue sur la donnée ne sont pas contradictoires, mais complémentaires. Les données sont réelles tant qu'elles résistent aux épreuves². Or, leur capacité à résister à ces épreuves dépend précisément de l'ensemble du réseau sociotechnique qui les fait tenir. La solidité de la donnée repose en effet sur sa capacité à être révisée et mise à jour pour s'ajuster aux différents usages³. Ainsi, contrairement à une perception courante, la solidité d'une donnée ne réside pas tant dans sa stabilité, son immuabilité, mais dans sa souplesse, sa capacité à répondre à des usages variés. La « boîte noire » de la donnée n'est jamais totalement refermée, la donnée évoluant régulièrement pour s'adapter à différents besoins. La donnée repose ainsi sur une « instabilité » contrôlée⁴. Cependant, dans le même temps, pour ces utilisateurs, la donnée est une boîte noire, un objet technique stabilisé sur lequel ils peuvent s'appuyer pour construire leurs propres données métiers.

Ainsi, ce qui donne aux données de référence un caractère réaliste – et qui les distingue alors des données métiers – est l'ampleur de l'investissement de forme sur lesquelles elles reposent. Comme le souligne Alain Desrosières, « *c'est l'ampleur de l'investissement de forme réalisé dans le passé qui conditionne la solidité, la durée et l'espace de validité des objets ainsi construits* »⁵. C'est parce que les producteurs investissent, notamment⁶, dans des

¹ A. Desrosières, « Discuter l'indiscutable. Raison statistique et espace public. », art cit.

² On reprend ici le principe mis en avant par Bruno Latour : « Est réel ce qui résiste dans l'épreuve ». B. Latour, *Pasteur, op. cit.*, p. 244.

³ Emmanuel Didier et al., « La solidité des institutions. Les statistiques de "victimation" de l'Insee (1996-2006) », *Genèses*, 2009, vol. 1, n° 74, p. 128-144.

⁴ Une analogie peut être tracée avec le mode de développement dans les communautés *open source*. Neff et Stark évoquent un état de brouillons permanents (« Permanently Beta ») pour ces logiciels soumis à des processus de révision permanents. Gina Neff et David Stark, « Permanently Beta: Responsive Organization in the Internet Era » dans Philip N. Howard et Steve Jones (eds.), *Society Online. The Internet in Context*, Thousand Oaks, Sage Publications, 2004, p. 173-188.

⁵ A. Desrosières, *La politique des grands nombres, op. cit.*, p. 19.

⁶ L'investissement réalisé se matérialise de différentes façons : il est en premier lieu financier avec un coût initial conséquent et un budget annuel dédié (250 K€ hors RH) ; il est humain avec un service dédié composé de 35 agents, il est technique avec des dispositifs visant à toujours plus de précision et d'exhaustivité ; il est enfin

dispositifs de traçabilité de l'usage de leurs données qu'ils sont capables de leur donner un statut de réalité : d'en faire des « objets qui tiennent *au triple sens de leur robustesse propre (résistance à la critique), de leur capacité à se combiner entre eux, et enfin de ce qu'ils tiennent les hommes entre eux en les incitant (ou parfois en les contraignant) à user de ce langage à visée universaliste* »¹.

B. Le SIG et l'émergence d'un agencement territorial

Ces propriétés des données de référence qui les distinguent des données métiers conduisent à des formes différentes d'attachement. Alors que les secondes attachent des directions métiers à une représentation spécifique de l'espace urbain, formant des agencements fonctionnels, les premières attachent des acteurs multiples et hétérogènes à une représentation commune de l'espace.

1) Le SIG : un objet-frontière

Les données de référence mettent en relation les services de la communauté urbaine appartenant à des mondes hétérogènes. Les directions de l'eau, de la voirie, de la propreté, du foncier et de l'urbanisme peuvent se coordonner au travers du langage commun que constituent les données de référence. Cette représentation partagée rend possible l'action en commun. Plus généralement, les données de référence fournissent un cadre commun de description de l'espace à l'ensemble des acteurs de l'action publique métropolitaine. En effet, à partir de 1995, par le biais de conventions d'échange, la communauté urbaine de Lyon s'inscrit dans une démarche volontariste de diffusion de ses données de référence à l'ensemble de ses « partenaires ». La gratuité de cette mise à disposition encourage la diffusion massive de ses données qui deviennent alors le référentiel commun à l'ensemble des acteurs de l'agglomération². En 2012, 52 communes sur les 58 que comptent alors le Grand Lyon ont signé ces conventions d'échanges³, ainsi que près d'une trentaine d'administrations

normatif avec la volonté de s'appuyer sur des conventions externes (normes, etc.) pour solidifier davantage les données.

¹ A. Desrosières, *Pour une sociologie historique de la quantification*, *op. cit.*, p. 12.

² En 2003, ces modalités de diffusion externe de données sont complétées par la vente de plans et de fichiers informatiques aux particuliers et aux institutions hors convention, et par la mise à disposition d'une solution de consultation interactive par internet du système urbain de référence. Dans ce service, appelé Géonet, les utilisateurs peuvent uniquement consulter les données.

³ Les communes n'ayant pas signé de convention d'échange sont des petites communes ne disposant pas de système d'information géographique.

diverses (Compagnie Générale des Eaux (devenue Veolia Eau), Chambre d'agriculture du Rhône, Direction Départementale de l'Équipement, Sytral, Direction Départementale de la Sécurité Publique du Rhône, SDIS, divers syndicats intercommunaux, etc.)¹.

Cette capacité des données de référence à mettre en relation des mondes sociaux aux intérêts hétérogènes a conduit plusieurs auteurs² à considérer les SIG comme des objets-frontières³. Ils sont en effet à la fois suffisamment flexibles pour pouvoir être utilisés dans des contextes d'action variés et suffisamment robustes⁴ pour fournir une représentation commune du territoire entre ces acteurs. Le caractère numérique des données à référence spatiale permet la jonction, la connexion entre les différents sous-systèmes techniques, applications et métiers de la communauté urbaine. En permettant à chaque direction de la communauté urbaine de Lyon de produire sa propre représentation de l'espace à partir de références communes, le SIG connecte autant qu'il sépare ces différents groupes. Les données de référence SUR assurent ainsi la coopération tout en préservant les usages spécifiques de chacun.

2) Un agencement territorial

Les données de référence font toutefois plus que mettre en relation des mondes hétérogènes autour d'une représentation partagée. Par les attachements qu'elles créent, elles participent à la construction des entités qui leur sont attachées. De fait, les données de référence donnent consistance parallèlement à l'institution communautaire et à son territoire⁵.

¹ Parallèlement, en 2010, Géonet est disponible pour 1500 utilisateurs internes au Grand Lyon, 53 communes et 7 partenaires du Grand Lyon (SDIS, OPAC du Rhône, OPAC du Grand Lyon, SYTRAL, Agence d'urbanisme, etc.).

² Francis Harvey et Nick Chrisman, « Boundary objects and the social construction of GIS technology », *Environment and Planning A*, 1998, vol. 30, n° 9, p. 1683-1694.

³ Le concept d'objet-frontière, introduit par Star et Griesemer à partir d'une étude ethnographique du travail scientifique, permet de penser le rôle des objets dans la coordination entre acteurs. Ils le définissent ainsi : *“Boundary objects are objects which are both plastic enough to adapt to local needs and the constraints of the several practices employing them, yet robust enough to maintain a common identity across sites. They are weakly structured in common use, and become strongly structured in individual-site use. These objects may be abstract or concrete. They have different meanings in different social worlds but their structure is common enough to more than one world to make them recognizable, a means of translation. The creation and management of boundary objects is a key process in developing and maintaining coherence across intersecting social worlds”*. Susan Leigh Star et James R. Griesemer, « Institutional Ecology, “Translations” and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology », *Social Studies of Science*, 1989, vol. 19, n° 3, p. 387-420.

⁴ Cette robustesse repose sur une infrastructure partagée dont la solidité dépend de l'investissement de forme réalisé, comme cela a été démontré précédemment.

⁵ Dans une perspective proche, Delphine Gardey analyse comment différentes matérialités ont donné consistance à la territorialité de l'Assemblée nationale, entendue comme un espace à distinguer du territoire de l'Etat, une
Antoine COURMONT – « Politiques des données urbaines » - Thèse IEP de Paris – 2016

Émerge alors un agencement territorial qui assure à la communauté urbaine de Lyon un nouveau pouvoir à l'échelle de l'agglomération.

A sa création, l'institution communautaire ne disposait pas d'un territoire propre, mais se composait des territoires associés des communes membres. Les plans et les données dont dispose alors la communauté urbaine proviennent principalement des communes membres de l'intercommunalité. Ils fournissent une représentation parcellaire de l'agglomération : une mosaïque de plans communaux dont la vision est très disparate selon les municipalités. A partir des années 90, les données de référence du SIG participent à la construction du territoire du Grand Lyon en le représentant comme un espace homogène délimité par des frontières. Grâce à cet outil, la communauté urbaine dispose désormais d'une représentation exhaustive de son territoire.

Ca a apporté que si je veux faire une construction, j'appelle le centre de données urbaines, qu'est-ce que vous avez sur cet endroit ? J'ai le plan de voirie au 200^e, j'ai les tracés des lignes EDF, des lignes de gaz, j'ai la position des bistrots, parce que si je veux faire un ensemble scolaire, il ne faut pas un bistrot à moins de 200 mètres. **Bref, j'ai tout... j'ai tout le territoire** à disposition de tout le monde en fonction de ses besoins¹.

La création du SIG participe à la constitution du territoire de la communauté urbaine en le rendant perceptible. Pour la première fois, « tout le territoire » est représenté de manière uniforme sur le périmètre des 55 communes qui constituent la communauté urbaine. Ainsi, les données de référence donnent consistance au territoire du Grand Lyon, tout comme la statistique américaine a donné consistance à l'espace national des Etats-Unis², ou que le seuil de pauvreté a créé la pauvreté, le quotient intellectuel l'intelligence et les sondages de Gallup l'opinion publique³.

hétérotopie qui assure à l'institution autonomie et souveraineté. Delphine Gardey, *Le Linge du Palais-Bourbon. Corps, matérialité et genre du politique à l'ère démocratique*, Lormont, Le Bord de l'Eau, 2015, 256 p.

¹ Source : Entretien de Jean Barboyon, entré à la Communauté urbaine en 1969, il est ingénieur en chef du centre de données urbaines du Département Développement de 1984 à 1990. Entretien disponible sur le site <http://www.40ans.grandlyon.com/?p=6415> (Consulté le 3 juin 2015).

² E. Didier, *En quoi consiste l'Amérique ? Les statistiques, le New Deal et la démocratie*, op. cit.

³ Voir respectivement : I. Hacking, « Façonner les gens : le seuil de pauvreté », art cit ; O. Martin, *La Mesure de l'esprit. Origines et développement de la psychométrie, 1900-1950*, op. cit. ; Loïc Blondiaux, *La Fabrique de l'opinion. Une histoire sociale des sondages*, Paris, Seuil, 1998, 600 p ;

Plus généralement, sur le « pouvoir créateur » des quantifications, voir Theodore Porter, *Trust in numbers: the pursuit of objectivity in science and public life*, s.l., Princeton University Press, 1995, 326 p.

Parallèlement, les données de référence donnent consistance à l'institution communautaire qui devient un « être collectif¹ » autonome doté d'un pouvoir d'action à l'échelle de l'agglomération. Les données de référence créent une interdépendance entre des mondes sociaux : elles attachent ensemble les différents « métiers » qui composent l'institution. Alors que les données métiers reflètent la mosaïque des services intervenants de manière autonome sur la ville, par les données de référence, ces « métiers » sont tenus de fonctionner ensemble. A ce titre, il est significatif que le SIG ait été mis en place lorsque l'institution intercommunale se voyait attribuer de larges compétences en matière d'urbanisme, qui nécessitait de passer de la juxtaposition des métiers à leur coordination. Cette création d'interdépendance entre les mondes sociaux produit de la centralité : un pouvoir qui contrôle et institue des règles de coexistence faisant valoir sa domination sur les acteurs agissant sur son territoire. Ce pouvoir agit de façon indirecte : les données de référence ne prescrivent pas des actions, mais transforment les cadres de représentation au travers desquels sont pensées et mises en œuvre ces actions.

Par l'intermédiaire des données de référence, la communauté urbaine de Lyon est devenu l'acteur central incontournable des politiques d'agglomération. La constitution du SIG marque un changement majeur de l'histoire institutionnelle de l'intercommunalité qui n'est plus considérée uniquement comme une organisation au service des communes, mais qui poursuit la mise en œuvre de politiques publiques à l'échelle de l'agglomération. Interrogé sur le changement majeur intervenu au cours des quarante premières années de la communauté urbaine, Jean-Pierre Bué ne déclare-t-il pas : « *De toute cette évolution extraordinaire, je retiendrais, sans doute plus que tout, la mise à disposition de tous des bases de données géographiques².* » La diffusion des données géographiques révèle l'affirmation du pouvoir du Grand Lyon comme acteur de référence de l'agglomération. Alors que la communauté urbaine de Lyon était contrainte d'utiliser les plans produits par les communes, l'agence d'urbanisme ou l'Etat, ce sont désormais ces acteurs qui sont contraints d'utiliser les données du Grand Lyon pour agir sur son territoire.

¹ Thomas Angeletti et Aurélien Berlan, « Les êtres collectifs en question », *Tracés*, 2015, vol. 3, n° 29, p. 7-22.

² Source : Entretien de Jean-Pierre Bué, cadres à la retraite de la Communauté urbaine. Entretien disponible sur le site <http://www.millenaire3.com/interview/regards-croises-sur-les-differents-mandats-et-leurs-realisation-de-1970-a-2009> (Consulté le 6 janvier 2016).

IV. Conclusion : données et territorialisation

A l'issue de ce parcours explorant les opérations de mise en données de la ville, différentes modalités d'attachement des données ont été mises en avant. En premier lieu, les données sont constituées de multiples attachements, formant un réseau sociotechnique, composé d'entités hétérogènes, humains et non-humains, actualisé en permanence pour maintenir la solidité de la chaîne de la donnée. En second lieu, les données « métiers » attachent les services communautaires à l'espace urbain, par le biais d'une convention, dont la définition repose sur l'usage qui sera fait de la donnée. Données et usages sont ainsi étroitement liés. Enfin, les données de référence au cœur du SIG attachent les différents acteurs de l'action publique autour d'une représentation commune et stabilisée de l'espace urbain. Par l'ampleur des investissements réalisés pour établir les conventions sur lesquelles elles reposent, ces données parviennent à être des instruments de coordination entre des mondes sociaux hétérogènes.

Pour conclure ce chapitre, je souhaite revenir sur la question posée en introduction : qu'est-ce que les données *font* à l'espace urbain ? Les données ne visent en effet pas tant à décrire objectivement l'espace urbain qu'à être capable d'agir sur lui et de le maîtriser, c'est à dire de faire advenir le territoire de la Métropole de Lyon. En le représentant, elles le font exister et perdurer de manière tangible. Le territoire n'existe que parce que l'on est capable de le représenter, de le mettre en données, de le matérialiser, de le rendre calculable comme l'a souligné Stuart Elden¹. En s'inscrivant dans une perspective foucauldienne², il démontre que le territoire est une forme d'organisation spatiale historiquement et géographiquement située, associée à l'émergence de l'Etat moderne occidental³ et au développement de techniques de mesure et de découpage de l'espace. Le territoire est selon lui une technologie politique ; il est

¹ Stuart Elden, *The Birth of Territory*, Chicago, University of Chicago Press, 2013.

² Constatant que Michel Foucault focalise son analyse sur la population plutôt que sur le territoire, Stuart Elden propose d'appliquer l'approche foucauldienne à la notion de territoire, d'autant que celui-ci reste peu conceptualisé en science politique. Stuart Elden, « Governmentality, Calculation, Territory », *Environment and Planning D: Society and Space*, 2007, vol. 25, n° 3, p. 562-580 ; Stuart Elden, « Land, terrain, territory », *Progress in Human Geography*, 2010, vol. 34, n° 6, p. 799-817.

³ Stuart Elden s'oppose en cela à Saskia Sassen qui propose une définition historique du territoire moins rigoureuse. Le territoire est selon elle un assemblage organisationnel composé d'un espace délimité, d'une autorité et de droits. Dissociée de l'Etat moderne, cette théorisation lui permet de distinguer trois formes d'assemblages territoriaux (l'Etat médiéval, l'Etat moderne et l'Etat global) pour analyser les recompositions territoriales liées à la mondialisation. Saskia Sassen, *Territory, Authority, Rights: From Medieval to Global Assemblages*, Princeton, Princeton University Press, 2006, 502 p.

l'espace pensé comme une catégorie politique, c'est-à-dire possédé, délimité, cartographié, calculé et contrôlé. Les instruments qui participent à la construction du territoire comme technologie politique sont nombreux : géométrie, arpentage, navigation, cartographie, statistiques, récits, fortifications, images, etc. Les données constituent une nouvelle technique de gouvernement de l'espace. Elles ne sont pas des instruments de représentation d'un territoire qui serait déjà-là et prêt à être saisi, elles participent à la construction, à la gestion et au maintien du territoire. Le suivi des opérations quotidiennes de mise en données de l'espace urbain révèle les processus de territorialisation¹ à l'œuvre au sein de la Métropole de Lyon. Si la production des données et des territoires est étroitement liée, cela implique de considérer les territoires *en train de se faire* dans leur pluralité et de manière dynamique.

La diversité des modèles de représentations de l'espace véhiculés par les données démontre que de multiples notions de territoire coexistent au sein d'une même institution. Le territoire de la direction de la voirie diverge de celui de la direction de la propreté, qui est encore différent de celui de la direction de l'eau. On ne peut ainsi envisager le territoire de manière univoque : il faut penser les territorialités² dans leur diversité et leur pluralisme. Elles

¹ La territorialisation est entendue ici comme l'ensemble des processus de construction des territoires. Le géographe Bernard Debarbieux en propose la définition suivante : « *l'ensemble des actions, des techniques et des dispositifs d'action et d'information qui façonnent la nature ou le sens d'un environnement matériel pour le conformer à un projet territorial* ».

Bernard Debarbieux, « Territoire-Territorialité-Territorialisation : aujourd'hui encore, et bien moins que demain... » dans Martin Vanier (ed.), *Territoires, Territorialité, Territorialisation. Controverses et perspectives*, Rennes, Presses Universitaires de Rennes, 2009, p. 75-89 ;

Cette signification s'éloigne de celle traditionnellement employée en science politique, où la territorialisation de l'action publique interroge les effets de territorialité sur les conditions de formulation des principes de l'action publique. La territorialisation n'est alors qu'une *contextualisation spatiale* de l'action publique. Les travaux à ce sujet pointent le fait que le territoire, plus que le secteur (agriculture, industrie, culture, etc.) serait devenu le principe de définition et de légitimation des politiques publiques.

Patrice Duran, « Territorialisation » dans Romain Pasquier, Sébastien Guigner et Alistair Cole (eds.), *Dictionnaire des politiques territoriales*, Paris, Presses de Sciences Po, 2011, ; Alain Faure et Anne-Cécile Douillet, *L'action publique et la question territoriale*, s.l., Presses Universitaires de Grenoble, 2005 ; Alain Faure et al. (eds.), *Action publique et changements d'échelles : les nouvelles focales du politique*, s.l., L'Harmattan, 2007.

² La territorialité est définie par Robert Sack comme « *the attempt by an individual or a group to affect, influence, or control people, phenomena, and relationships, by delimiting and asserting control over a geographic area. This area will be called the territory* ». Claude Raffestin a critiqué cette définition qui réduit la territorialité à un contrôle sur une aire en faisant « *l'économie du pouvoir par assimilation de la territorialité humaine à celle animale, privée de symbolisation et donc de codification* ». Il propose ainsi de ne pas réduire la territorialité à l'appropriation matérielle d'un espace, mais de prendre en considération l'identification culturelle qui lui est associée. Michel Lussault précise encore cette définition avec la notion de modèle territorial, c'est-à-dire la « *représentation de l'organisation légitime du territoire, sur laquelle s'appuie l'action d'un acteur politique* ». Selon lui, cette idéologie territoriale est nécessaire « *pour faire « tenir ensemble » les différents composants de leur espace d'action, pour lier solidement entre elles des unités discrètes – des lieux, des aires – et par là contribuer à produire la continuité nécessaire à l'existence du territoire légitime* ». Robert David Antoine COURMONT – « *Politiques des données urbaines* » - Thèse IEP de Paris – 2016

ne sont pas similaires d'un espace à un autre, d'une institution à une autre, puisqu'elles dépendent des pratiques et des projets territoriaux qui sont assignés à l'espace. Sur un même espace, plusieurs territoires coexistent, s'entrecroisent, se superposent, peuvent entrer en conflit, etc. Une même institution politique peut d'ailleurs porter plusieurs projets d'organisation spatiale en fonction des secteurs sur lesquels elle intervient, même si ceux-ci sont souvent englobés dans une totalisation territoriale (« le territoire de l'institution »). Cette complexité territoriale invite à penser les territoires dans leurs relations en analysant ainsi les configurations territoriales¹ à l'œuvre sur un espace donné.

Pour analyser le rôle des données dans la constitution des territoires, j'ai eu recours à la notion d'agencement. Développé par la sociologie de l'acteur-réseau, et notamment par Michel Callon², la notion d'agencement constitue le dispositif à partir duquel l'action parvient ou non à se maintenir. En suivant les différentes modalités d'attachement des données, ce chapitre met en évidence deux formes principales d'agencement spatial qui s'articule au sein de la Métropole de Lyon.

- Un agencement fonctionnel qui émerge de l'attachement entre un service communautaire (« un métier ») et une représentation de l'espace urbain à visée opérationnelle. Les données répondent à la nécessité de disposer d'une représentation du périmètre spatial et thématique – qui peut déborder le périmètre administratif - sur lequel une direction doit agir.
- Un agencement territorial qui fait tenir ensemble des acteurs hétérogènes autour d'une représentation commune de l'espace urbain. Les données de référence donnent alors consistance à l'institution métropolitaine et à son territoire lui conférant un nouveau pouvoir à l'échelle de l'agglomération.

Sack, *Human territoriality: its theory and history*, New York, Cambridge University Press, 1986 ; Claude Raffestin, « Espace et pouvoir » dans A Bailly (ed.), *Les concepts de la géographie humaine*, Paris, Masson, 1995, ; Michel Lussault, *L'Homme spatial : La construction sociale de l'espace humain*, Paris, Seuil, 2007, p. 114 ; Michel Lussault, « Territorial (Modèle) » dans Jacques Lévy et Michel Lussault (eds.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris, Belin, 2003, p. 917.

¹ Le terme de configuration territoriale renvoie à une conception relationnelle des territoires. Il invite à focaliser l'analyse sur les modalités de disposition des territoires entre eux et le champ de leurs relations possibles.

² La notion d'agencement a en particulier été développée par Michel Callon dans le prolongement de la critique deleuzienne de la notion foucauldienne de dispositif.

M. Callon, Y. Millo et F. Muniesa, *Market Devices, op. cit.*, p. 2.

Chapitre 2

Mettre en circulation : Généalogie de la politique d'ouverture des données de la communauté urbaine de Lyon

La politique d'ouverture de données de la communauté urbaine de Lyon est le résultat de la conjonction de plusieurs initiatives, au travers desquelles se croisent des acteurs variés, aux intérêts pluriels et aux valeurs divergentes. Clément, chef de projet *open data*, m'en fait une longue présentation lors de mon arrivée au sein du Grand Lyon, qui en souligne la complexité¹.

Le projet *open data* est une initiative de la Direction des systèmes d'information. Il s'intègre dans une conjonction de projets. La diffusion des données géographiques avec Inspire, la centrale de mobilité portée par la Direction de la Voirie, et le projet Ville intelligente de la Direction au développement économique. [...] Il y a aussi eu un effet de mode *open data*. [...] Aujourd'hui la Direction au développement économique s'est appropriée le projet *open data*, car [le projet] Smart City n'a toujours rien produit, et le Living Lab est encore loin d'être ouvert. Donc, l'ouverture des données publiques, ça leur permet de dire, on avance sur le sujet, on a fait quelque chose. Et ce, alors que le bilan de fin de mandat approche. Pour la Direction des systèmes d'information et son directeur, l'ouverture des données publiques est une opportunité pour le Système d'information de prendre une place plus stratégique qu'elle n'a aujourd'hui au sein du Grand Lyon. [...] Aujourd'hui, la Délégation au développement urbain ne comprend pas l'intérêt de l'*open data*, ou alors ils le comprennent mais ils n'ont aucune envie d'y participer. Pour eux, il s'agit d'un projet Inspire, d'ouverture des données géographiques, donc, pas de leur compétence. Le projet d'ouverture des données est aujourd'hui perçu par la majorité des services comme un truc de geeks. Ils n'y voient aucun intérêt pour le moment. Il faut dire qu'il ne s'agit pas d'une commande politique, et qu'aujourd'hui, c'est l'approche technologique qui a été privilégiée. Donc, c'est mal vendu. Il y a un gros travail à faire là dessus. C'est à nous de les mobiliser, de leur vendre le projet en s'adaptant à leur cœur de métier. Par exemple, on est content, on a réussi à vendre l'*open data* à la déléguée au développement durable, en lui présentant le sujet comme une opportunité pour l'environnement, pour qu'elle soit prête à porter et à défendre le sujet. Laura et Meryl, de la direction de la prospective, sont aussi intéressées par le sujet de l'*open data* dans une démarche de participation... Mais bon, pour le moment, ils n'avancent pas trop sur le sujet car il n'y a pas de portage politique et de demande sur cette problématique de la participation. Après, il y a les opposants au projet dès l'origine comme Harald. Lui, il a toujours été contre l'*open data*. Mais, c'est un opposant intelligent. Il a un argumentaire sérieux derrière, de défense des politiques publiques et de défense d'une réalité économique locale, avec une problématique de développement économique territorial et de défense des petits acteurs face aux gros comme Google, etc.. Après, ça pose la question, est-ce que la collectivité doit contrôler les politiques publiques qui sont menées à l'extérieur de la collectivité ?²

L'*open data* est tour à tour associé au développement économique, à la *smart city*, à la directive Inspire, à la Centrale de mobilité, à la « mode *open data* », à des annonces politiques, à la participation citoyenne et la transparence, à une volonté stratégique de la part de certains acteurs de repositionner leur service au sein de l'institution, etc. Cette diversité

¹ Je m'en étonne d'ailleurs dans mon journal de terrain : « Commentaires et premières impressions : Les démarches autour du suivi de projet semblent hyper compliquées. Il y a beaucoup de procédures, aux sigles étranges, à respecter. Beaucoup d'acteurs interviennent dans le processus, souvent juste pour faire approuver une décision. D'où, un nombre très important de réunions en tout genre pour présenter, discuter et faire accepter les différentes orientations des projets. Dès lors, il semble compréhensible qu'un projet tel que l'ouverture des données soit difficile à faire accepter au sein d'une institution telle que le Grand Lyon : le nombre d'acteurs à mobiliser, à convaincre, à « intéresser » est très élevé. Et surtout, chaque acteur semble pouvoir (et vouloir ?) mettre son grain de sel pour que le projet s'adapte à ses intérêts. »

² Source : Journal de terrain – Réunion avec Clément (23 novembre 2012)

d'enjeux fait intervenir de multiples acteurs qui ne sont pas moins hétérogènes : une élue aux nouvelles technologies, des entreprises, des « geeks » du service informatique, un « opposant » à l'*open data*, des personnes qui y sont indifférentes, des directeurs de service, des agents provenant de différents services de la communauté urbaine, des acteurs intéressés, d'autres à intéresser, etc. Au final, des données, on n'en trouve que peu de traces explicites. Pourtant, elles sont présentes en filigrane de ce récit et prennent des aspects multiples : informations géographiques pour les communes, ressources économiques pour les entreprises, prises pour la participation citoyenne, etc.

Ce chapitre tente d'y voir plus clair dans ce projet d'ouverture de données en retraçant la « mise en politique » de la question de la circulation des données publiques. Comment les données « fermées » ont-elles été constituées en problème devant être résolu ? Comment la mise en circulation des données s'est imposée comme un objet d'action publique à part entière ? Pour répondre à ces questions, je suis au plus près la trajectoire de ce problème public de son émergence à sa mise sur l'agenda de la communauté urbaine, en passant par les nombreuses bifurcations qu'il subit entre temps. Cette perspective généalogique¹ souligne que la politique d'ouverture des données n'a pas une seule origine linéaire, mais qu'elle est le résultat d'une constellation de valeurs, de visions et de pratiques plurielles qui conduisent différents acteurs à parler d'ouverture de données de manière contingente. La carrière de l'*open data* nous mènera d'arènes en arènes de problématisation dans lesquelles différents acteurs interviennent, et au cours desquelles les cadrages du problème seront l'objet de multiples interprétations avant de se stabiliser².

L'objet de ce chapitre est le processus de problématisation de la politique d'ouverture des données de la communauté urbaine. Il se caractérise par une succession d'épreuves³ lors desquelles la nature du réseau et du problème est transformée. *« L'épreuve est un moment au cours duquel les personnes font preuve de leurs compétences soit pour agir, soit pour désigner, qualifier, juger ou justifier quelque chose ou quelqu'un : un être. C'est donc le*

¹ La démarche généalogique est associée aux travaux de Michel Foucault. Elle est une interrogation historique des conditions de possibilité que les choses soient ce qu'elles sont, en montrant les instabilités, les incertitudes et les lignes de fracture qui bousculent la linéarité apparente des choses.

² Dans la continuité des travaux de Howard Becker et Erving Goffman, le concept de « carrière » est utilisé couramment pour suivre l'étude des politiques publiques. Cf. Daniel Cefaï, « La construction des problèmes publics. Définitions de situations dans des arènes publiques. », *Réseaux*, 1996, vol. 14, n° 75, p. 43-66.

³ Pour une présentation détaillée de la notion d'épreuve, voir supra Chapitre 3.

moment d'incertitude par excellence puisqu'il est l'occasion de s'accorder sur l'état des êtres, qui peuvent être des personnes et/ou des choses »¹. Les épreuves sont ainsi « *l'ensemble des situations et des dispositifs qui réalisent et font tenir le problème* »² de l'ouverture des données. Au travers de ces épreuves, un processus de traduction se met en œuvre à l'issue duquel les acteurs s'accordent sur une vision commune du problème à traiter. Lors de ces moments, à l'issue incertaine, ce ne sont pas seulement les cadres de perception de l'*open data* qui évoluent, mais l'ensemble des actants de la situation. La problématisation est une phase de formulation du problème et des acteurs concernés³ : certains sont durablement associés, ou, a contrario, dissociés de la politique publique en construction.

Ce chapitre se divise en trois parties qui constituent autant d'étapes du processus de problématisation de l'ouverture des données de la communauté urbaine de Lyon. Tout d'abord, trois arènes indépendantes au sein desquelles la diffusion des données publiques est évoquée sont présentées : la transposition de la directive européenne INSPIRE, un projet de Centrale de mobilité et le mouvement citoyen de l'*open data*. Ensuite, la constitution d'une « équipe projet » marque une nouvelle phase de problématisation visant à associer dans un même projet ces diverses volontés de diffusion de données. La stratégie d'ouverture de données de la communauté urbaine est alors élaborée sous l'influence des directions de la voirie et du développement économique. Enfin, l'instrumentation de cette action publique au travers des licences constitue une solidification de cette stratégie de diffusion qui éloigne le Grand Lyon des principes du mouvement de l'*open data*⁴. Avant d'entrer plus en détail, une précision lexicale s'impose : dans ce chapitre, deux significations du terme d'*open data* coexistent. Dans sa signification courante, l'*open data* est synonyme d'ouverture ou de diffusion des données publiques. Toutefois, dans une signification restreinte, il désigne un mouvement citoyen visant à la mise à disposition des données publiques selon certains principes stricts.

¹ Mohamed Nachi, *Introduction à la sociologie pragmatique*, Paris, Armand Colin, 2006, p. 57.

² Benjamin Lemoine, *Les valeurs de la dette. L'Etat à l'épreuve de la dette publique*, Ecole nationale supérieure des mines de Paris, Paris, 2011, p. 24.

³ Cette conception de la problématisation s'inspire du processus de traduction présenté par Michel Callon dans son article fondateur sur la domestication des coquilles Saint-Jacques. Cf. Michel Callon, « Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques dans la Baie de Saint-Brieuc », *L'Année sociologique*, 1986, n° 36.

⁴ Dans ce chapitre, deux significations du terme d'*open data* coexistent. Dans sa signification courante, l'*open data* est synonyme d'ouverture ou de diffusion des données publiques. Toutefois, dans une signification restreinte, il désigne un mouvement citoyen visant à la mise à disposition des données publiques selon certains principes stricts.

I. Trois arènes indépendantes de problématisation de la diffusion de données publiques

Le récit de la problématisation de l'ouverture des données de la communauté urbaine de Lyon débute au sein de trois arènes respectivement centrées autour de l'échange de données géographiques, de l'*open data* et de la mise en commun des informations sur la mobilité urbaine. Des acteurs multiples y échangent des perceptions variées de ce que doit être la diffusion des données et ce à quoi celle-ci doit répondre. Dans la lignée de l'héritage pragmatiste de John Dewey¹ et de Joseph Gusfield², on prend alors conscience que le problème public n'est pas tant un fait qu'une action collective qui le fait advenir. Celle-ci se déroule au sein d'arènes³ publiques, des espaces d'interactions au travers desquelles s'échangent, se confrontent, se coordonnent différentes perceptions du monde social⁴. L'arène n'existe pas a priori, elle se déploie autour d'une situation problématique, sans que l'on puisse lui assigner des frontières déjà instituées. La publicité de ces arènes est variable : ces interactions peuvent se dérouler tout autant dans des espaces publics que des espaces confinés⁵.

Le suivi de ces échanges souligne les jeux de cadrages et de re-cadrages du problème public qui sont à l'œuvre au sein de ces arènes⁶. En effet, comme le souligne Daniel Cefaï,

¹ John Dewey s'est intéressé au processus de publicisation, c'est-à-dire à la constitution d'un public envisagé comme la mobilisation d'acteurs affectés par un problème social. John Dewey, *Le public et ses problèmes*, Paris, Gallimard, 2010.

² Joseph Gusfield, *La culture des problèmes publics : L'alcool au volant : la production d'un ordre symbolique*, Paris, Economica, 2008, 354 p.

³ Le concept d'arène, largement développé par Daniel Cefaï, doit être distingué de ceux de marché, où se rencontrent des intérêts individuels issus stratégies rationnelles, de champ, qui réduit la nature des problèmes sociaux à l'état des rapports de force et des conflits d'intérêts entre classes sociales, et d'agora, un lieu de délibération et d'échange d'arguments qui fait émerger la raison publique. L'arène doit plutôt être considérée comme un « milieu » où se croisent des champs d'expérience et des univers de discours autour d'*issues* particulières. Cf. Daniel Cefaï, « Qu'est-ce qu'une arène publique ? Quelques pistes pour une approche pragmatiste » dans Daniel Cefaï et Isaac Joseph (eds.), *L'Héritage du pragmatisme. Conflits d'urbanité et épreuves de civisme*, La Tour d'Aigues, Editions de l'Aube, 2002, p. 51-82.

⁴ D. Cefaï, « La construction des problèmes publics. », art cit.

⁵ Claude Gilbert, Emmanuel Henry et Isabelle Bourdeaux, « Lire l'action publique au prisme des processus de définition des problèmes » dans Claude Gilbert et Emmanuel Henry (eds.), *Comment se construisent les problèmes de santé publique*, Paris, La Découverte, 2009, p. 7-33 ; Yannick Barthe, *Le pouvoir d'indécision : la mise en politique des déchets nucléaires*, Paris, Economica, 2006, 239 p.

⁶ La notion de cadrage fait suite aux travaux d'Erving Goffman, qui met en avant la multiplicité des « cadres » au travers desquels chacun fait l'expérience d'une même situation. Cf. Erving Goffman, *Les cadres de l'expérience*, Paris, Les Editions de Minuit, 1991, 576 p.

« nous n'avons jamais affaire qu'à une diversité de versions du problème public, qui correspondent à la diversité de ses modes de construction »¹. Le problème de l'ouverture des données est ainsi saisi de façon multiple selon les acteurs qui s'en emparent. Alors que pour l'Union européenne, il vise à faire émerger une représentation européenne de l'environnement, les agents du service de l'information géographique de la communauté urbaine de Lyon n'y voient qu'un moyen de moderniser leurs modes de travail, et les militants de l'*open data* rien de moins qu'une modalité pour renouveler les relations gouvernants/gouvernés. Au fil des interactions, ces cadrages initiaux seront l'objet de recadrages qui feront émerger d'autres versions du problème public et qui finiront par le stabiliser.

Ces cadrages sont autant de « bricolages cognitifs » qui permettent de rapprocher un problème, une solution et une politique publique². La problématisation ne se réduit pas en effet à la « formulation générale d'un problème », mais comprend « l'identification des voies à suivre pour le traiter, et par conséquent la désignation des acteurs les plus légitimes pour le prendre en charge »³. Au cours de chacune de ces problématizations de la politique d'ouverture des données, le problème, l'instrument, les données et le public ciblé sont catégorisés différemment.

A. Arène 1 : La traduction communautaire de la directive européenne INSPIRE

Le projet d'une plateforme de diffusion de données de la communauté urbaine fait son apparition au sein du service de l'information géographique. Il trouve son origine dans la directive Inspire, portée par l'Union européenne qui impose aux administrations publiques des pays membres la mise à disposition de leurs données géographiques. Cette directive a été traduite au sein de la communauté urbaine de Lyon pour s'inscrire dans une volonté de modernisation de la diffusion de ses données aux communes membres de la structure intercommunale. Ce projet est porté par le service de l'Information géographique (Igé) appartenant à la direction des systèmes d'information (DSIT). La problématisation de cette

¹ D. Cefaï, « La construction des problèmes publics. », art cit.

² Philippe Zittoun, *La fabrique politique des politiques publiques. Une approche pragmatique de l'action publique.*, Paris, Presses de Sciences Po, 2013, p. 140.

³ Yannick Barthe, « Le recours au politique ou la problématisation politique « par défaut » » dans Jacques Lagroye (ed.), *La politisation*, Paris, Belin, 2003, p. 475-492.

politique publique assigne à la communauté urbaine une fonction traditionnelle de fourniture de services aux communes.

1) La diffusion des données géographiques au Grand Lyon

Depuis le milieu des années 90, la communauté urbaine de Lyon diffuse des données issues de son système d'information géographique (SIG) auprès des communes et de ses partenaires. Ce partage de données poursuit deux objectifs. En premier lieu, il s'agit de fournir des instruments de représentation du territoire aux acteurs n'ayant pas de ressources propres. En second lieu, cette diffusion vise à coordonner les connaissances de chacun et à partager un référentiel commun afin de renforcer la cohérence entre les différents acteurs agissant sur le même territoire. Il s'agit donc de se doter d'un instrument commun d'objectivation de l'agglomération nécessaire aux discussions collectives.

Une unité du service géomatique se consacre spécifiquement à la diffusion de ces informations géographiques. Ces données sont diffusées dans le cadre de conventions d'échange, par la fourniture, soit de plans papier, soit de fichiers numériques. La directive européenne INSPIRE va contribuer à renouveler profondément ces modalités de diffusion.

2) La directive européenne INSPIRE impose aux administrations publiques la diffusion de leurs informations géographiques

Afin d'encourager l'intégration des exigences de protection de l'environnement dans la mise en œuvre des politiques publiques, l'Union européenne adopte en 2007 la directive INSPIRE¹ qui impose aux administrations publiques la diffusion de leurs données géographiques. L'UE souhaite favoriser le partage d'informations entre administrations afin de constituer une représentation de l'environnement européen par agrégation des données obtenues aux niveaux inférieurs. Pour cela, cette directive vise à établir une infrastructure d'information géographique² (IDG), c'est-à-dire un instrument de mise en réseau des bases de

¹ Directive 2007/2/CE

² Une infrastructure d'information géographique (IDG, en anglais Spatial Data Infrastructure, SDI) se compose « des métadonnées, des séries de données géographiques et des services de données géographiques ; des services et des technologies en réseau ; des accords sur le partage, l'accès et l'utilisation ; et des mécanismes, des processus et des procédures de coordination et de suivi établis, exploités ou mis à disposition » (Directive 2007/2/CE, transposé dans la législation française par l'ordonnance n°2010-1232 du 21 octobre 2010).

données géographiques produites par les acteurs publics européens¹. Afin de standardiser les modalités d'échange de données, la directive INSPIRE, oblige les administrations des États membres à fournir des données selon des règles de mise en œuvre communes, à constituer des catalogues de données respectant les normes et standards d'interopérabilité et à proposer des services en ligne d'accès aux données et aux métadonnées². Les données concernées sont de trois types : les données nécessaires au repérage sur le territoire³, les données générales complémentaires⁴ et les données thématiques⁵.

L'originalité de l'instrument IDG est de produire une européanisation des représentations de l'environnement tout en préservant les multiples représentations localisées de l'espace européen. Conformément au principe de subsidiarité, les administrations locales, plus proches des phénomènes sociaux, sont jugées les plus légitimes pour produire, gérer et maintenir les données. Mais, les techniques d'échanges automatisés de données permettent aux autorités centrales d'agréger et de réagencer plus aisément ces données selon leurs propres besoins. Par la mise en place de cet instrument commun d'objectivation, l'Union européenne légitime en effet un domaine d'intervention à l'échelle européenne. *« La non centralisation de ces patrimoines de données sur l'environnement européen offre aux autorités européennes la possibilité d'organiser un travail de redéfinition implicite de ce qu'est « l'environnement européen ». Ainsi, la définition européenne de l'environnement diffère sensiblement de celle donnée par différents États membres, ce qui conduit INSPIRE à mettre en réseaux des juridictions nationales (voire infra-nationales) qui ne sont parfois pas directement en charge des questions environnementales⁶. »* La directive INSPIRE est

¹ La directive INSPIRE n'est pas la première réglementation européenne en matière de diffusion de données publiques. La directive 2003/98/CE, qui la précède, fixe les règles et les conditions de réutilisations des données publiques. S'inscrivant dans les programmes de l'Union européenne sur la société de l'information, elle s'inscrit dans une logique de réutilisation économique de ces informations par les entreprises en vue de produire de nouveaux services.

² Les services de recherche et de consultation doivent être mis à disposition gratuitement. La question du coût de téléchargement et de transformation des données est laissée à la libre disposition de chaque administration.

³ Les données définies dans l'annexe 1 : Référentiels de coordonnées, réseaux de transport, hydrographie, adresses et nom de lieux, parcelles cadastrales, sites protégés, unités administratives...

⁴ Les données définies dans l'annexe 2 : Altimétrie, occupation des terres, géologie, ortho-imagerie, géologie.

⁵ Les données définies dans l'annexe 3 : Bâtiments, vocation des sols, santé et sécurité des personnes, installations industrielles et agricoles, données sur l'environnement, démographie, zones à risque industriel, conditions atmosphériques...

⁶ Matthieu Noucher, « Infrastructures de données géographiques et flux d'information environnementale », *Netcom*, 2013, vol. 27, n° 1-2, p. 120-147.

caractéristique des instruments de gouvernance de l'Union européenne¹ qui s'appuient sur des techniques de transferts de connaissance et d'objectivation² pour européaniser les politiques publiques des Etats membres³.

La directive INSPIRE révèle une première problématisation de la diffusion des données géographiques. Elle met en évidence l'existence d'un enjeu à traiter – la politique environnementale européenne, d'une cause – la dispersion des données entre administrations – et d'une solution – la mise en place d'une infrastructure de données géographiques. La directive INSPIRE articule ainsi dans un même mouvement un instrument résolvant une cause et solutionnant un problème public, tout en y associant des acteurs nombreux et divers - l'ensemble des autorités publiques productrices de données - et en définissant des règles de mise en œuvre.

3) La traduction de la directive INSPIRE par la communauté urbaine de Lyon

Cette nouvelle réglementation européenne oblige la communauté urbaine de Lyon à renouveler les modalités de diffusion de ses données pour mettre en place une infrastructure d'information géographique. En effet, la fourniture annuelle de données par le biais de plan imprimé ou de CD d'actualisation ne répond pas aux préconisations de la directive. Pour étudier les modalités d'application de celle-ci, le service de l'information géographique du Grand Lyon va réaliser une étude qui donne lieu à la rédaction d'un schéma directeur sur les données géographiques⁴. Adopté en 2010, ce nouveau schéma directeur est consacré, pour une large part, à la diffusion des données géographiques. Il répond aux demandes des communes et des partenaires qui réclament davantage de données – métiers et de référence, accessibles plus aisément par un « guichet unique ».

¹ Bastien Sibille, « Voir l'Europe pour la faire. Un système d'informations géographiques dans la "gouvernance" européenne. », *Politique européenne*, 2010, n° 31, p. 147-172 ; M. Noucher, « Infrastructures de données géographiques et flux d'information environnementale », art cit.

² Comme par exemple le benchmarking. Isabelle Bruno, « Y a-t-il un pilote dans l'Union ? », *Politix*, 2008, n° 82, p. 95-117.

³ Sabine Saurugger et Yves Surel, « L'eupéanisation comme processus de transfert de politique publique. », *Revue internationale de politique comparée*, 2006, vol. 13, n° 2, p. 179-211.

⁴ Il s'agit du troisième schéma directeur sur les données géographiques du Grand Lyon après celui de 1992 portant sur les données urbaines au sens large et celui de 2005 portant sur les données géographiques de référence.

Le schéma directeur de 2010 joint ces demandes aux exigences de la directive INSPIRE. La mise en place d'une infrastructure de données spatiales permet en effet un accès facile, rapide et standardisé aux données, tout en assurant les quatre services de base requis par la directive Inspire (découverte, visualisation, transformation, téléchargement). Le schéma directeur de 2010 trace ainsi les grandes lignes des modalités d'ouverture des données géographiques, qui se matérialise par la mise en œuvre, à l'automne 2012, d'une plateforme de catalogage et de téléchargement de ces données¹.

La mise à disposition des données du Grand Lyon sur une infrastructure numérique ne constitue pas une rupture, mais s'inscrit dans la continuité de la démarche de diffusion de l'information géographique. Depuis sa création, le service de l'information géographique poursuit une politique de partage des données du SIG à l'ensemble des acteurs œuvrant sur le territoire de la métropole lyonnaise (services, communes, partenaires). La plateforme *catalogue.data.grandlyon.com* amplifie la diffusion en l'étendant à un public indifférencié et en facilitant l'accès aux données².

La directive INSPIRE est ainsi traduite par la communauté urbaine de Lyon pour s'inscrire dans sa politique de diffusion de données géographiques aux communes de l'intercommunalité. L'enjeu pour le Grand Lyon n'est pas tant l'absence de représentation de l'environnement à l'échelle européenne que l'amélioration de la coordination entre les acteurs des politiques communautaires. La cause de ce problème n'est plus l'éparpillement des données géographiques, mais les difficultés d'accès immédiat aux données actualisées. La résolution passe, par contre, par un instrument similaire : un catalogue centralisé de données géographiques communautaires offrant des outils de visualisation et de téléchargement. La mise en place d'une plateforme permettant l'échange de données ne vise alors pas seulement à répondre aux contraintes réglementaires européennes, mais à s'inscrire dans la continuité de services fournis aux communes membres. L'ouverture des données doit permettre d'améliorer les modalités de diffusion de l'information géographique entre les différents acteurs agissant sur le territoire (services, communes, partenaires). Décloisonner la donnée, ici restreinte à l'information géographique, entre les autorités publiques, vise à faciliter la conception, le

¹ Accessible sur Internet à l'adresse : <http://catalogue.data.grandlyon.com> Cette plateforme remplit l'ensemble des conditions imposées par la directive INSPIRE. Les données sont disponibles gratuitement après validation d'une licence de type ODbL.

² Les réutilisateurs n'ont plus l'obligation de signer une convention d'échange de données, de souscrire au service *Géonet* ou d'acheter les données auprès du Grand Lyon.

suivi et l'évaluation des politiques publiques. S'ils font référence à la directive INSPIRE, les documents de présentation de cette démarche ne mentionnent jamais sa dimension environnementale. Le « public » visé n'est pas l'Union européenne, mais les communes et autres utilisateurs des données géographiques de la communauté urbaine. Elle correspond ainsi à une des facettes traditionnelles de l'intercommunalité : une institution au service des 58 communes qui la composent.

B. Arène 2 – La Centrale de mobilité : diffuser les données pour répondre aux enjeux environnementaux, économiques et de mobilité

La deuxième arène de problématisation de la diffusion des données publiques émerge au sein du service mobilité urbaine de la direction de la voirie de la communauté urbaine de Lyon. La mise à disposition des données y est perçue comme une solution aux situations de congestion routière et un remède à la pollution environnementale. Elle doit permettre le développement de services d'informations multimodales par des acteurs privés, services qui favoriseront le report modal des automobilistes vers des modes doux (transports en commun, vélo, marche).

1) La Centrale de Mobilité : la diffusion des données au cœur de la politique de mobilité

Pour résoudre des problématiques de mobilité, la direction de la voirie propose à la fin des années 2000 une nouvelle stratégie basée sur la diffusion de données de transport et le développement de services d'informations aux voyageurs. Cette stratégie trouve son origine dans un projet de R&D, intitulé Mobiville, coordonné par le cabinet de conseil Algoé, et auquel participaient les communautés urbaines de Lyon et de Toulouse. Il visait à mettre en place des services d'information en temps réel pour répondre à une perspective de mobilité durable selon le postulat que « *l'utilisation des transports publics (bus, tramway, métro, vélos en libre service) peut être simplifiée et plus efficace grâce à une information voyageur à forte valeur ajoutée*¹. » Ce projet était porté par Harald², alors manager au sein de l'entreprise

¹ Dossier de synthèse de l'évaluation du projet Mobiville (Février 2010)

² Ingénieur de formation, Harald rejoint la communauté urbaine de Lyon en 2010 pour prendre en charge la coordination des programmes de gestion de circulation et de transports en commun. Il a travaillé auparavant durant 20 ans chez Algoé, une société de conseil en management, pour lequel il menait des missions relatives à la mobilité urbaine. Il a notamment participé au programme de navigateur urbain sur Lyon Mobiville. Il est également expert auprès de la Commission européenne sur la mobilité urbaine et les systèmes de transport intelligent (ITS).

Algoé. Il s'achève au début de l'année 2010, et, quelques mois plus tard, Harald rejoint la communauté urbaine de Lyon pour prendre en charge la coordination des programmes de gestion de circulation et de transports en commun. Il élabore alors une stratégie d'information voyageur et de mise à disposition des données mobilité en se basant sur les conclusions de l'évaluation du projet Mobiville.

Le problème posé est le suivant : face aux enjeux environnementaux et aux risques de saturation des infrastructures routières, les acteurs publics souhaitent encourager le report modal de la voiture individuelle vers les transports en commun et les modes doux (vélo, marche, etc.). Plusieurs solutions ont été apportées pour répondre à ce défi. La première a été de développer l'offre de transports en commun par la construction de nouvelles infrastructures. La deuxième a été réduire l'espace de voirie attribué à la voiture¹. La nouvelle stratégie de la direction de la voirie apporte une troisième réponse visant à agir sur la demande (le comportement des voyageurs) par le développement de services d'information sur les offres de transport disponible sur l'agglomération. Le postulat sous-jacent est qu'un accès facilité à l'information orientera le comportement des voyageurs qui se tourneront vers des solutions alternatives à l'usage de la voiture. *« Les services d'information sur la mobilité sont de puissants outils permettant une meilleure gestion de son déplacement, de connecter concrètement les réseaux entre eux et, enfin, d'agir sur les comportements des usagers². »* Le problème est que si ces informations sur les solutions de mobilité sont multiples sur le territoire du Grand Lyon, elles sont fragmentées et non coordonnées. Plusieurs sites d'information coexistent en effet selon les modes : InfoTrafic pour le réseau routier et les sites TCL et Multitud pour les transports en commun.

Face à cette situation problématique, le service mobilité de la direction de la voirie défend une « vision à 2020 » : *« passer d'une situation où l'information est éclatée entre les modes, n'est pas en temps réel pour l'essentiel, n'est accessible que depuis un PC ou via les radios partenaires pour l'information trafic, n'est disponible qu'une partie du temps, à une information déplacements en temps réel (fiable), tous modes (multimodale), tout le temps et partout (embarquée et GPS), pour tous (ciblée et au moindre coût) par tous (partagée et interactive)³. »* Pour réaliser ce dessein, le service mobilité préconise la mise en place d'une

¹ Ces deux solutions se retrouvent conjointement dans la mise en place des lignes de tramway.

² Source : Dossier de presse Conseil Communautaire, (27 juin 2011)

³ Source : Présentation DG Cab (30 mai 2011)

« Centrale de mobilité¹ ». Celle-ci doit « rassembler en un seul lieu l'ensemble des informations concernant les services de déplacements aujourd'hui très éclatés, et développer des services innovants d'information sur les déplacements pour les habitants de la Communauté urbaine »². Un des axes de cette stratégie repose sur la mise à disposition de ces données de déplacement auprès des acteurs privés « qui en font la demande (comme, par exemple, les opérateurs internet d'informations déplacements : Infomobile, ViaLeChemin, Vtrafic, etc.) pour permettre la création de services nouveaux »³. L'objectif est de multiplier les canaux de diffusion de l'information tout en s'assurant que l'utilisation des données soit cohérente avec les politiques publiques de mobilité. Approuvée par le conseil communautaire du 27 juin 2011, qui valide son budget de 760 000 euros (dont 600 000 à la charge de la communauté urbaine), cette nouvelle stratégie inscrit la diffusion des données au cœur de la politique de mobilité de la communauté urbaine.

2) Le projet Optimod : la diffusion des données mobilité au service de l'innovation et du développement économique

A la recherche de financements complémentaires, la direction de la voirie va traduire ce projet de Centrale de mobilité pour l'inscrire dans une perspective d'innovation et de développement économique. La communauté urbaine de Lyon dépose en effet un dossier de subvention auprès de l'Ademe, dans le cadre du programme « Investissements d'avenir » qui vise à favoriser « l'innovation et la création d'emplois non délocalisables » tout en prenant en compte la transition écologique et énergétique⁴. La Centrale de mobilité devient alors le projet Optimod (Optimiser la mobilité durable) dont l'enjeu n'est plus uniquement d'encourager le report modal, mais également de favoriser le développement économique.

Coordonné par la communauté urbaine de Lyon, un partenariat public-privé de R&D associant 13 acteurs est mis en place⁵. Il doit contribuer au développement économique des

¹ Les Centrales de mobilité s'inscrivent dans la continuité des services d'information multimodale, obligatoires pour les AOT des agglomérations de plus de 100 000 habitants par la loi SRU du 13 décembre 2000 (loi n°2000-1208).

² Source : Délibération 2011-2279 du conseil communautaire (27 juin 2011)

³ Source : Délibération 2011-2279 du conseil communautaire (27 juin 2011)

⁴ Source : <http://www.ademe.fr/entreprises-monde-agricole/innover-developper/investissements-davenir> (Consulté le 9 juillet 2015)

⁵ Les entreprises IBM, Cityway, Parkeon, Renault Trucks, Orange Business Service, Autoroutes Trafic, Geoloc Systems, Phoenix, la ville de Lyon, et les laboratoires LIRIS, CETE de l'Est et le LET.

entreprises partenaires : pour sept millions d'euros engagés, 83 millions de retombées économiques sont espérées. Le soutien aux entreprises françaises est explicite : « *C'est un véritable challenge fonctionnel, technique, et ergonomique, qui devrait permettre à la plus grosse PME du secteur en France, CityWay, de rivaliser avec ses homologues allemands (Hacon, MDV) qui aujourd'hui déploient leur savoir faire dans différents pays de l'Union Européenne et en dehors de l'Union (USA notamment)*¹. » Le partenariat mis en place par la communauté urbaine de Lyon doit faciliter l'export des PME françaises sur les marchés européens et internationaux. Le recours à l'appel à subvention de l'ADEME contribue ainsi à traduire le projet initial pour associer des enjeux de développement économique à cette politique d'information mobilité.

Au centre de ce projet, figure un « entrepôt » de données, qui centralise, consolide, puis diffuse « l'ensemble des données mobilité du territoire » : topologie des réseaux, temps théorique, temps réel, temps prévu à une heure, qualité de l'air, etc. Le service mobilité fait en effet le constat que des informations existent sur quasiment tous les modes de transport, mais que l'information est éclatée en autant de sites que d'organisations. « *L'éclatement institutionnel sur les réseaux de transports (Sytral, Région, SNCF, CG, Etat, Autoroutiers, ...etc) se reflète aujourd'hui dans l'information à l'utilisateur. La centrale de mobilité vise à remédier à cela en connectant les réseaux entre eux. Il s'agira donc d'aller au-delà d'une simple agrégation des services existants mais de les dépasser dans une optique centrée sur l'utilisateur qui est un utilisateur de tous les modes*². »

Ce projet d'ouverture des données diffère ainsi de celui élaboré initialement par le service d'information géographique dans la mesure où il vise à rassembler des données éparses, puis à les diffuser à des acteurs externes. Surtout, l'objectif n'est pas tant de partager une même représentation du territoire, que de construire une nouvelle représentation incluant l'ensemble des offres de transport de l'agglomération auparavant dispersées. L'enjeu n'est plus de moderniser l'administration, mais d'insérer la diffusion des données dans des politiques publiques portées par la communauté urbaine. La diffusion des données s'inscrit

¹ Source : Dossier de presse « Signature Convention Optimod » (5 avril 2012)

² Source : Cahier des charges – Définition d'une stratégie opérationnelle et des outils de communication sur l'information sur les déplacements au Grand Lyon (Janvier 2012)

ainsi à la fois dans une politique environnementale, de mobilité et de développement économique¹ !

C. Arène 3 – Le mouvement de l’open data

Une troisième arène de problématisation de la mise à disposition des données publiques émerge parallèlement au deux précédentes et s’impose au Grand Lyon : le mouvement de l’*open data*. Originaire des Etats-Unis, ce mouvement, porté des acteurs de la société civile, se traduit pas la mise en place de portails *open data* par des collectivités territoriales françaises. L’agglomération lyonnaise n’est pas en reste puisqu’elle voit la création de l’organisation OpenData69 qui tente de faire pression sur la communauté urbaine de Lyon. Face à cette effervescence d’initiatives sur le sujet, la vice-présidente aux nouvelles technologies annonce la mise en place d’une plateforme *open data* lors d’une conférence publique, qui marque l’entrée du Grand Lyon dans le mouvement de l’*open data*.

1) Les origines anglo-saxonnes du mouvement de l’open data

En 2006, le journal anglais The Guardian lance une campagne de transparence administrative intitulée « *Give us back our crown jewels* » appelant à une libération des données publiques, qui, financées par l’impôt des contribuables, devraient leur être librement accessibles. Cette mobilisation est considérée comme le moment initiatique du mouvement « open data »². Elle fait suite à la création de l’association *Open knowledge foundation*³ (OKFN) en 2004 par l’économiste Rufus Pollock, dont les travaux de recherche pointent la valeur économique des données ouvertes⁴. L’OKFN milite en faveur de la « culture libre »

¹ Alors que traditionnellement, les politiques de déplacements urbains se caractérisent par une tension entre enjeux économiques et environnementaux. « *La mise à l’agenda des déplacements urbains s’opère généralement à travers une tension entre préoccupations économiques et environnementales. [...] Le registre économique exclut le registre environnemental et réciproquement selon les circonstances et les opportunités dont peuvent bénéficier les groupes sociaux qui soutiennent ces deux logiques* ». Bernard Jouve, *Les politiques de déplacements urbains en Europe. L’innovation en question dans cinq villes européennes.*, Paris, L’Harmattan, 2003, p. 22.

² Il faut préciser que l’*open data* était depuis bien longtemps une revendication au sein du milieu académique, dont certains membres réclamaient l’ouverture des données de la recherche afin de faciliter les collaborations.

³ L’Open Knowledge Foundation est une association à but non lucratif qui promeut la culture libre (open knowledge, open access, open science, open data). Elle a été créée au Royaume-Uni en 2004. L’organisation s’est ensuite internationalisée, elle compte aujourd’hui des déclinaisons locales dans près de 60 pays.

⁴ A partir d’une étude de la branche locale de l’OKFN, Stefan Baack étudie le lien entre les mouvements *open source* et *open data* établis par cette association. Stefan Baack, « Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism », *Big Data & Society*, 2015, vol. 2, n° 2.

(open knowledge, open access, open science, open data) dont elle propose une définition en 2005. « *Open data and content can be freely used, modified, and shared by anyone for any purpose* »¹.

Parallèlement, un mouvement similaire en faveur de l'ouverture des données publiques émerge aux Etats-Unis sous l'impulsion d'organisations de la société civile (Sunlight Foundation², Code for America³) avec une foi et une exigence dans la transparence des administrations⁴. La conférence de Sébastopol, organisée en décembre 2007, marque l'acte fondateur de ce mouvement. A l'initiative de Tim O'Reilly et de Carl Malamud, elle réunit une trentaine de militants des mouvements du logiciel et de la culture libres⁵. Aux termes de deux jours de réunions, ils formalisent l'*open government data* en définissant huit principes :

1. *Data Must Be Complete*
2. *Data Must Be Primary*
3. *Data Must Be Timely*
4. *Data Must Be Accessible*
5. *Data Must Be Machine processable*
6. *Access Must Be Non-Discriminatory*
7. *Data Formats Must Be Non-Proprietary*

¹ Source : <http://opendefinition.org/> (Consulté le 10 juillet 2015)

² La Sunlight Foundation est une organisation non-gouvernementale américaine, créée en 2006, qui défend les principes du « gouvernement ouvert ». Son objectif est de développer la transparence et l'*accountability* des administrations publiques. La devise de l'organisation, empruntée à Louis Brandeis, « *La lumière du soleil est le meilleur des désinfectants* », s'inscrit dans une rhétorique de la lumière qui éclairerait et soignerait des administrations opaques.

³ Code for America est une organisation non gouvernementale américaine, fondée en 2009, dont le but est de développer l'usage des technologies ouvertes par les administrations publiques afin de favoriser leur efficacité et de promouvoir la participation. Elle se revendique du « civic hacking » : des développeurs engagés pour promouvoir les innovations civiques. Particulièrement engagée auprès des administrations locales, l'organisation est divisée en « brigades » présentes dans plus d'une centaine de villes aux Etats-Unis, et également en Europe (Allemagne, Pologne) et au Japon.

⁴ Ce mouvement s'inscrit dans la culture du numérique qui promeut la libre circulation de l'information comme valeur fondamentale d'organisation de la société.

Fred Turner, *Aux sources de l'utopie numérique : De la contre-culture à la cyberculture*, Stewart Brand, un homme d'influence, Caen, C&F éditions, 2013, 430 p ; Sébastien Broca, *Utopie du logiciel libre. Du bricolage informatique à la réinvention sociale*, Paris, Le Passager clandestin, 2013, 282 p ; Benjamin Loveluck, « Internet, une société contre l'Etat ? Libéralisme informationnel et économies politiques de l'auto-organisation en régime numérique », *Réseaux*, 2015, vol. 4, n° 192, p. 235-270.

⁵ Parmi lesquels l'avocat et professeur de droit à Stanford Lawrence Lessig, le militant Aaron Swartz, le fondateur du site FixMyStreet Tom Steinberg, le créateur du service d'informations locales EveryBlock Adrian Holovaty, le développeur du site GovTrack.us Joshua Tauberer.

8. *Data Must Be License-free*¹

Ces huit principes sont devenus la définition canonique de l'*open data*. Ils s'inscrivent dans la tradition du mouvement open source en adoptant une interprétation restrictive de l'adjectif « ouvert ». La donnée doit être brute (non retraitée), accessible dans un format non-propriétaire, sous une licence permettant son usage et sa diffusion sans discrimination et à un coût marginal. Cette catégorisation modifie la problématisation associée aux données dans la mesure où elle autonomise les données de leurs usages ultérieurs. Les données ouvertes doivent pouvoir être utilisées pour une variété large d'usages, par un public indifférencié, sans aucun contrôle de la part des producteurs publics.

Pour les militants de l'*open data*, ces principes sont nécessaires pour que l'ouverture des données publiques atteigne ses ambitions, à savoir l'amélioration de la démocratie en favorisant la transparence de l'administration et la participation. Le mouvement de l'*open data* va alors rapidement se développer tant à l'échelle des Etats que des villes².

Encadré 3 : Open data & Open government

La demande sociale en faveur de l'*open data* trouve écho lors de la campagne présidentielle de 2008 auprès du candidat Barack Obama qui s'engage en faveur de l'*open government*. L'*open government* est une idéologie selon laquelle il serait possible « d'ouvrir » le gouvernement par le biais de l'information³. En ayant accès à celle-ci, les citoyens pourraient contrôler et prendre part aux décisions gouvernementales, tout en mettant en œuvre des services à partir de ces données⁴. Au lendemain de son investiture, Barack Obama signe

¹ Source : <http://www.opengovdata.org/home/8principles> Consulté le 14 août 2013.

² Pour une analyse plus exhaustive de la généalogie du mouvement *open data* et ses liens avec le secteur de la réutilisation des informations publiques (PSI Re-use), voir Jonathan Gray, « Towards a Genealogy of Open Data », Glasgow, 2014 ; Pour une analyse de l'appropriation et de la traduction de l'*open data* par le gouvernement anglais, voir Jo Bates, « The Domestication of Open Government Data Advocacy in the United Kingdom: A Neo-Gramscian Analysis », *Policy & Internet*, 2013, vol. 5, n° 1, p. 118-137.

³ Pour une analyse critique de l'idéologie « open » appliquée au gouvernement, voir Harlan Yu et David Robinson, « The New Ambiguity of "Open Government" », *UCLA Law Review Discourse*, 2012, vol. 59, n° 178, p. 180-208 ; Nathaniel Tkacz, « From open source to open government : a critique of open politics », *Ephemera: Theory and Politics in Organization*, 2012, vol. 12, n° 4, p. 386-405.

⁴ A ce sujet, deux postures cohabitent. Certains estiment que les acteurs publics ne doivent pas avoir le monopole de la possession des données publiques, tandis que d'autres appellent à un retrait de l'Etat au profit d'acteurs privés, ces derniers étant jugés comme les mieux placés pour développer des services à partir des données publiques. David Robinson et al., « Government Data and the Invisible Hand », *Yale Journal of Law & Antoine COURMONT – « Politiques des données urbaines » - Thèse IEP de Paris – 2016*

trois mémorandums sur la transparence et le gouvernement ouvert, dans lesquels il demande aux administrations étatiques de respecter les principes de transparence, de participation et de collaboration. L'*open data* y figure en bonne place et se manifestera par le lancement du portail *data.gov* le 21 mai 2009. Dans le sillage des Etats-Unis, l'appropriation gouvernementale de ce mouvement de l'*open data* va ensuite se répandre dans le monde. De nombreux pays créent successivement des sites de mise à disposition de leurs données : le Royaume-Uni, la Nouvelle-Zélande, la Norvège, l'Australie, le Canada, le Maroc, le Kenya, le Chili, l'Italie, l'Espagne, les Pays-Bas, le Brésil, la France, etc. Sous l'impulsion des Etats-Unis, 65 pays vont se réunir au sein de l'*Open Government Partnership* afin de promouvoir les principes du gouvernement ouvert. En 2013, les chefs d'État du G8 signent une charte en faveur de l'ouverture des données publiques.

Encadré 4 : Les villes américaines s'emparent de l'open data. L'exemple de NYC

Parallèlement à l'administration centrale, plusieurs villes américaines vont s'emparer du sujet de l'*open data*. A New York, la politique d'ouverture de données trouve son origine dans l'organisation, en 2009, d'un concours d'applications, NYC BigApps, financé par le *New York City Economic Development Corporation*¹. Plusieurs agences de la ville mettent alors à disposition certaines de leurs données. Puis, en janvier 2010, l'entreprise publique chargée de la gestion des transports publics dans l'agglomération de New York, la *Metropolitan Transportation Authority*, publie l'ensemble de ses données horaires afin de favoriser le développement d'applications de services d'informations aux voyageurs qu'elle ne peut développer faute de ressources². En octobre 2011, la ville de New York lance son premier portail *open data*, suivi en mars 2012 d'une loi par laquelle le maire de New York, Bloomberg, impose la publication par défaut de l'ensemble des données des administrations de la ville³. Cette loi marque une nouvelle étape dans la politique d'ouverture de données,

Technology, 2009, vol. 11, n° 1 ; Tim O'Reilly, « Government as a Platform », *innovations*, 2010, vol. 6, n° 1, p. 13-40.

¹ Le *New York City Economic Development Corporation* est une société à but non lucratif chargée de favoriser le développement économique dans les cinq boroughs de la ville.

² "By making access to our data directly from our website, we are encouraging the developer community to do the work we can't to create apps that benefit our customers at no cost to the MTA.", Jay Walder, MTA Chair (13 janvier 2010)

Source : <http://www.mta.info/press-release/mta-headquarters/mta-launches-new-website> (Consulté le 11 août 2015)

³ Local law 11 of 2012, http://www.nyc.gov/html/doi/html/open/local_law_11_2012.shtml (Consulté le 11 août 2015)

puisqu'elle change le paradigme de l'open data : la mise à disposition des données devient la norme, la rétention des données par les services, l'exception. L'ouverture des données y est perçue comme une opportunité de rendre le gouvernement plus transparent et plus efficient, de favoriser la coordination entre administrations, d'encourager la participation du public et de créer des opportunités économiques.

2) La réception de l'open data en France

Transférées de pays en pays, les politiques d'*open data* font l'objet de traductions locales pour être adaptées à chaque environnement national. En France, le mouvement de l'open data émerge tout d'abord à l'échelon municipal sous l'impulsion de collectivités pionnières et de mouvements citoyens. Les politiques d'ouverture des données se diffusent ensuite sur l'ensemble du territoire avec la création de groupes citoyens locaux et la sensibilisation des acteurs publics. Si la question de la transparence administrative est ancienne¹, l'*open data* marque le passage d'une législation basée sur l'accès aux documents administratifs à une législation basée sur leur libre réutilisation dans une perspective de libéralisme informationnel².

La ville de Rennes est la première à s'emparer de la question de l'ouverture des données. A l'initiative de l'opérateur de transport public Keolis, un site open data mettant à disposition les données de vélo en libre service (VLS) est lancé en mars 2010. Des développeurs sont invités à s'emparer des données pour créer des solutions innovantes d'information à destination des usagers. Le site sera enrichi par les données de la collectivité, qui organisera, sur le modèle des villes américaines, un concours d'applications afin d'encourager la réutilisation des données³. La ville de Paris lui emboîtera le pas en lançant en janvier 2011 le site opendata.paris.fr, puis suivront les collectivités de Montpellier, Toulouse, Bordeaux, Nantes, de Saône et Loire, de Provence Alpes Côte d'Azur, etc.

¹ La Déclaration des droits de l'homme et du citoyen de 1789 stipulait déjà que « la société a le droit de demander compte à tout agent public de son administration » (article 15). En 1978, la loi CADA (Commission d'Accès aux Documents Administratifs) instaure un droit d'accès à l'information publique.

² Sur la notion de libéralisme informationnel, voir B. Loveluck, « Internet, une société contre l'Etat ? Libéralisme informationnel et économies politiques de l'auto-organisation en régime numérique », art cit.

³ 43 applications seront créées à l'issue de ce concours, dont une large majorité sur la thématique du transport et de la mobilité urbaine.

Encadré 5 : L'Union européenne : une logique de valorisation économique des données publiques

Quelques années avant l'émergence du mouvement *open data*, l'Union européenne s'est intéressée au potentiel économique de la réutilisation des données publiques. La Commission européenne multiplie les études qui évaluent les retombées économiques annuelles à plusieurs dizaines de milliards d'euros¹. En 2003, une directive sur la réutilisation des informations du secteur public est votée². Elle ne prend pas en compte l'accès des citoyens à l'information et se concentre exclusivement sur les aspects économiques en encourageant les Etats membres à faciliter la réutilisation de leurs données publiques³. La directive sera légèrement amendée en 2013 afin de s'aligner avec les principes du mouvement *open data*.

Encadré 6 : L'open data de l'État français : d'une approche patrimoniale aux données comme bien commun

A l'échelle nationale, la mission Etalab est créée en 2011 avec pour objectif la création d'un portail unique interministériel destiné à rassembler et à mettre à disposition l'ensemble des informations publiques de l'Etat⁴. La création de cette mission, placée sous l'autorité du Premier ministre, marque un tournant dans la gestion et la diffusion des données de l'Etat. En effet, celles-ci étaient auparavant l'objet d'une politique patrimoniale mise en place par l'Agence pour le patrimoine immatériel de l'Etat (APIE)⁵. Les données publiques étaient considérées comme la matière première d'un marché essentiel pour le secteur privé qui

¹ Les estimations sont très variables, de 27 milliards à 200 milliards selon les études. Cf. Graham Vickery, « Review of recent studies on PSI re-use and related market developments. », *Information Economics*, 2011, p. 41.

² Directive 2003/98/EC, dite 'PSI Directive'

³ La directive Inspire sur les informations géographiques peut également être envisagée sous l'angle des réutilisations économiques alors que se développent les industries des données géographiques. Katleen Janssen, *The Availability of Spatial and Environmental Data in the EU. At the Crossroads between Public and Economic Interests*, Alphen a/d Rijn, Kluwer Law International, 2010, 656 p.

⁴ Le Rapport Riester « Amélioration de la relation à l'utilisateur » de février 2010 préconisait la création « d'une plateforme d'innovation de services « Etat Lab » permettant aux acteurs tiers de développer des services innovants à partir des données publiques ». L'objectif était alors de copier le modèle de l'Appstore d'Apple et de produire un « Appstore gouvernemental ».

⁵ Créée en 2007 suite au rapport de la commission Lévy-Jouyet sur l'économie de l'immatériel, l'APIE est rattachée au Ministère de l'Economie et des Finances. Elle a pour mission de valoriser les actifs immatériels de l'Etat (images, marques, lieux publics, savoir-faire, propriété intellectuelle, données, etc.).

pouvait en faire un usage commercial. L'APIE est chargée de valoriser cette ressource¹ en favorisant l'accès aux données publiques par la création d'un portail unique² et en proposant aux administrations des licences encadrant leur diffusion³. Disposant de missions similaires⁴, la mission Etalab semble à première vue un doublon de l'APIE. Elle marque pourtant une évolution dans la définition des informations publiques, en s'éloignant de cette logique de marchandisation des données publiques pour les considérer comme des vecteurs de transparence et d'innovation dont l'utilisation doit être « libre, facile et gratuite »⁵. Cette logique se matérialise par l'adoption du principe de gratuité par défaut des données publiques⁶, la publication de la « Licence ouverte », compatible avec les licences libres existantes, et le lancement du portail data.gouv.fr en décembre 2011. La création de la mission Etalab manifeste une reprise en main des conditions de diffusion des informations publiques par les services du Premier ministre au détriment du ministère de l'économie. Elle marque également une nouvelle problématisation de cette politique publique en l'associant aux objectifs politiques du mouvement de l'open data. L'Etat transforme sa politique publique de diffusion des informations publiques d'une logique patrimoniale vers une ouverture large,

¹ L'APIE commande un rapport universitaire à des économistes afin d'étudier les potentialités de valorisation des informations du secteur public. Julien Penin et al., *La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale*, Strasbourg, BETA (Bureau d'Economie Théorique et Appliquée), 2010.

² Le plan France Numérique 2012 pour le développement de l'économie numérique, publié en octobre 2008, indique ainsi : « Au titre de ses missions, l'APIE pilotera la conception du projet de portail unique d'accès, en étroite collaboration avec le SGG et la DGME ; dans le souci de garantir l'adéquation avec les besoins des jeunes pousses de l'économie numérique, le STSI sera associé à la conception de ce site. Action n°39 : Favoriser le développement de nouveaux produits et services en créant un portail unique d'accès aux données publiques dont la conception sera pilotée par l'APIE. » L'APIE ne mettra cependant jamais en place ce portail unique de diffusion des données publiques.

³ L'APIE propose deux licences types aux administrations souhaitant mettre à disposition leurs données : la « Licence de réutilisation d'informations publiques délivrée en application de la loi n°78-753 du 17 juillet 1978 et prévoyant une livraison unique des informations » et la « Licence de réutilisation d'informations publiques délivrée en application de la loi n°78-753 du 17 juillet 1978 et prévoyant une livraison successive des informations ».

⁴ Le décret n°2011-194 créant la mission Etalab stipule en effet « La mission « Etalab » est chargée de la création d'un portail unique interministériel destiné à rassembler et à mettre à disposition librement l'ensemble des informations publiques de l'Etat, de ses établissements publics administratifs et, si elles le souhaitent, des collectivités territoriales et des personnes de droit public ou de droit privé chargées d'une mission de service public. » (article 2) et « La mission « Etalab » coordonne l'action des administrations de l'Etat et apporte son appui à ses établissements publics administratifs pour faciliter la réutilisation la plus large possible de leurs informations publiques. » (article 3).

⁵ Circulaire du 26 mai 2011 relative à la création du portail unique des informations publiques de l'Etat « data.gouv.fr » par la mission « Etalab » et l'application des dispositions régissant le droit de réutilisation des informations publiques.

⁶ Décret n° 2011-577 du 26 mai 2011 relatif à la réutilisation des informations publiques détenues par l'Etat et ses établissements publics administratifs

redéfinissant en cela la donnée qui cesse d'être un actif stratégique pour devenir un bien acquérant de la valeur par sa circulation.

Les partisans du mouvement de l'open data associent l'ouverture des données publiques à l'innovation numérique et à un renouveau de la relation entre gouvernants et gouvernés. Ces deux objectifs, qui peuvent sembler très éloignés l'un de l'autre, sont pourtant très souvent combinés dans les discours justifiant ces nouvelles politiques de diffusion de données publiques¹. Ces politiques publiques répondraient ainsi d'une part à un contexte de développement de l'économie numérique présentée comme une nouvelle révolution industrielle dont les données seraient le nouveau pétrole, et, d'autre part, à une situation de défiance des citoyens envers les institutions publiques, laquelle serait solutionnée par les données définies comme un vecteur de transparence administrative². Enfin, la politique d'open data est associée à la modernisation de l'action de l'Etat³. Le partage des données publiques doit également améliorer le fonctionnement administratif et permettre une meilleure efficacité de l'action étatique en développant de nouvelles formes de coordination.

¹ En témoigne par exemple, le préambule de la circulaire du 26 mai 2011 : « *Le développement de l'économie numérique et de l'innovation technologique constitue un enjeu majeur tant en termes de croissance et d'emplois, que de compétitivité et d'accès à l'information. En accédant librement aux informations publiques dont disposent les administrations, la communauté des développeurs et des entrepreneurs peut dès lors être en mesure de créer de nouveaux usages et des services applicatifs innovants. En matière d'innovation technologique, l'offre crée souvent la demande. En mettant à disposition ses informations publiques, l'Etat participe à la construction de la société numérique. Cette stratégie d'ouverture des données publiques (« Open Data ») illustre l'ambition de la politique industrielle et d'innovation du Gouvernement.*

La réutilisation libre, facile et gratuite des informations publiques est un levier essentiel pour favoriser la dynamique d'innovation qui sera portée par la communauté des développeurs et des entrepreneurs à partir des données mises en ligne sur « data.gouv.fr ». La créativité des développeurs et des entrepreneurs ne saurait se heurter à des cloisons artificielles qui ont trop souvent constitué des freins au développement de l'innovation dans notre pays. Le portail « data.gouv.fr » illustre ainsi l'engagement de l'Etat en faveur du renforcement de la compétitivité des entreprises françaises, qu'il s'agisse d'entrepreneurs individuels ou de petites, moyennes ou grandes sociétés.

« Data.gouv.fr » proposera des services en ligne afin de renforcer la transparence de la vie publique et la confiance des citoyens dans les institutions de la République. Ces services mettront en valeur le travail des administrations, contribueront à la transparence de l'action de l'Etat et éclaireront le débat public. Ils enrichiront ainsi la vie de notre démocratie.

Il convient donc que le portail « data.gouv.fr » mette à disposition librement, facilement et gratuitement le plus grand nombre d'informations publiques. La politique gouvernementale d'ouverture de ces informations doit être lisible et offrir à tous les réutilisateurs la sécurité juridique nécessaire au plein exercice de leur droit. »

² La mission Etalab place l'open data dans la continuité de la Déclaration des droits de l'homme et du citoyen de 1789 dont l'article 15 stipule « *La société a le droit de demander compte à tout agent public de son administration* ».

³ Le 31 octobre 2012, la mission Etalab est ainsi rattachée au Secrétariat général pour la modernisation de l'action publique (SGMAP), nouvellement créé par le Premier ministre Jean-Marc Ayrault.

Comme aux Etats-Unis, l'évolution de la politique de diffusion des données publiques est largement influencée par différents mouvements de la société civile qui assignent aux données ouvertes ces enjeux politiques. De nombreuses conférences et journées d'études sont organisées sur ce sujet contribuant à sensibiliser les acteurs publics, d'autant que des médias s'en font le relais (Owni, InternetActu, la Gazette des communes, etc.). A l'échelle locale dans un premier temps, puis à l'échelle nationale, plusieurs associations jouent un rôle important de lobbying et d'accompagnement des administrations dans la mise en œuvre des politiques d'open data. Travaillant depuis plusieurs années avec les collectivités, la Fing¹ lance en 2010 une campagne sur l'ouverture des données publiques en lien avec la ville de Rennes et la communauté urbaine de Bordeaux. Elle organise de nombreux événements et anime la réflexion sur le sujet de l'*open data*. En 2011, la Fing publie un guide pratique de l'ouverture des données publiques territoriales qui sera largement diffusé au sein des collectivités. Auparavant, l'association Regards citoyens², qui milite pour l'ouverture des données parlementaires, a travaillé conjointement avec les services de la mairie de Paris pour traduire et adapter la licence ODbL au droit français. Elle effectue également un travail de « lobbying citoyen » lorsque les projets d'open data portés par les collectivités ne respectent pas les principes de l'*open definition*³. L'association nantaise Libertic⁴ est une figure centrale du mouvement français de l'open data, par sa participation à de nombreux événements, ses discussions avec les collectivités, sa veille et ses réflexions structurantes publiées à intervalles réguliers sur son blog. Enfin, quelques jeunes entreprises du secteur de l'économie numérique (DataPublica, Captain Dash, OpenDataSoft, etc.) se font également les porte-voix de ce mouvement de l'open data, en mettant en avant le potentiel de développement économique qu'ils peuvent tirer de la réutilisation des données publiques. Elles contribuent à donner aux

¹ La Fondation Internet nouvelle génération (Fing) est une association française créée en 2000. Elle se définit comme un *think tank* sur les transformations numériques. Elle accompagne les entreprises, les institutions et les territoires à anticiper les mutations liées aux technologies et à leurs usages.

² Regards Citoyens est une association française créée en 2009. Face aux difficultés d'accès aux informations parlementaires, elle s'engage dans le mouvement de l'ouverture des données. Son objectif est de « proposer un accès simplifié au fonctionnement de nos institutions démocratiques à partir des informations publiques. » L'association se compose d'une dizaine de personnes, dont une majorité a une expertise en science informatique.

³ Notamment par des prises de position argumentées sous forme de billets de blogs. Cf. par exemple : <http://www.regardscitoyens.org/lopendata-cest-bien-plus-que-de-la-communication-politique> (Consulté le 13 juillet 2015).

⁴ Libertic est une association française créée en 2009 à Nantes. Elle a pour objectif de « promouvoir l'ouverture des données publiques, l'e-démocratie, le gouvernement 2.0 et d'accompagner notre territoire dans le développement et l'utilisation d'applications numériques d'utilité publique ». L'association est très engagée dans le mouvement de l'open data à l'échelle nationale et participe à de nombreux événements sur ce sujet. Elle est membre fondateur du collectif OpenDataFrance et coorganisateur de la première semaine européenne de l'*open data*.

administrations un élément tangible du potentiel de réutilisation économique des données publiques.

Ainsi, la mise à disposition des données publiques peut recouvrir des significations variées. Si la question de la transparence administrative est ancienne¹, le sujet a évolué avec le passage d'une législation basée sur l'accès aux documents administratifs à une législation basée sur leur réutilisation. Surtout, un mouvement en faveur de l'*open data* a émergé aux Etats-Unis, avant de se répandre dans le monde, dans le sillage de l'*open source* et de l'*open government*. En France, ce sont des collectivités locales qui ont investi tout d'abord ce sujet avant que l'Etat transforme sa politique publique de diffusion des informations publiques d'une logique patrimoniale vers une ouverture large, redéfinissant en cela la donnée qui cesse d'être un actif stratégique pour devenir un bien commun devant être partagé.

3) Le Grand Lyon à la « mode » de l'*open data* ?

Contrairement aux collectivités françaises pionnières sur le sujet, la communauté urbaine de Lyon ne s'inscrit pas, dans un premier temps, dans le mouvement de l'*open data*. Au sein de l'administration, certains agents suivent à distance les initiatives des municipalités rennaises ou parisiennes, mais aucun acteur ne s'empare de cette cause². Le sujet est pourtant évoqué comme un des 8 axes de transformation de l'administration électronique au Grand Lyon dans un rapport commandé en 2011 par la direction générale. Cependant, ce rapport ne sera jamais appliqué suite au changement de directeur général comme l'indique Clément, chef de projet *open data* :

[Le précédent DGA], quand il a été nommé, il nous a demandé [à Nicolas et Clément] une étude sur l'administration électronique. Il venait d'en faire faire une au Conseil Général de Seine Saint-Denis où il était en poste. On commande une étude sur le sujet à [un cabinet de conseil]. Ils nous ont produit un rapport, qui n'était pas très bon, mais dedans, ils préconisaient l'ouverture et la coproduction des données, la production de services... Donc, l'*open data* était déjà présent. Mais le projet a accouché d'une souris. Trois mois plus tard, le DGA est parti, et il faut avouer que ni Nicolas ni moi n'avions envie de promouvoir le sujet. On avait d'autres projets que l'on estimait plus importants³.

¹ La Déclaration des droits de l'homme et du citoyen de 1789 stipulait déjà que « la société a le droit de demander compte à tout agent public de son administration » (article 15). En 1978, la loi CADA (Commission d'Accès aux Documents Administratifs) instaure un droit d'accès à l'information publique.

² Ces agents ont connaissance des initiatives, notamment par la lecture des rapports de la FING ou d'autres revues professionnelles, mais ne s'inscrivent pas dans ce mouvement.

³ Source : Journal de terrain – Réunion avec Clément (23 novembre 2012)

Préconisé dans un rapport sur l'administration électronique, l'*open data* n'est pas mis à l'agenda en l'absence d'entrepreneurs de cause s'en emparant au sein de l'administration. Le DGA quitte la collectivité, le rapport commandé sombre dans l'oubli tout comme sa mesure en faveur de l'*open data*.

Toutefois, une coalition de cause, composée d'acteurs externes à l'institution, tente de s'emparer du sujet et de le mettre à l'agenda de la collectivité. Plusieurs revendications émergent de la société civile. En novembre 2011, le mouvement Open Data 69 est créé par plusieurs acteurs du numérique lyonnais (chercheurs, développeurs, entrepreneurs, élu municipal, etc.). Ce collectif d'activistes locaux œuvre pour la mise à disposition des données publiques sur la métropole lyonnaise. Leur mobilisation trouve un écho médiatique par la parution dans le journal local Rue 89 Lyon d'un article questionnant l'absence de projets d'*open data* au sein de l'agglomération lyonnaise¹.

Cette mobilisation d'acteurs de la société civile, et, surtout, l'adoption de cette cause médiatique par un nombre croissant de collectivités françaises pionnières va conduire la communauté urbaine de Lyon à leur emboîter le pas. Par une déclaration de sa vice-présidente aux nouvelles technologies, le Grand Lyon va s'inscrire dans cette « mode » de l'*open data* comme le relate un chef de projet.

A l'origine de la plateforme de diffusion de données, il y a aussi eu un effet de mode open data. A un moment, on avait des conférences sur le sujet toutes les semaines. Lors d'une de ces conférences, au salon des Interconnectés, en décembre 2011, la vice-présidente aux nouvelles technologies, Karine Dognin-Sauze annonce que le Grand Lyon va lancer sa propre plateforme d'ouverture des données dans les 18 mois. J'étais dans la salle, je suis resté scotché sur mon siège. Le sujet n'avait jamais été évoqué par le politique auparavant. Il a alors bien fallu s'y mettre².

Cette déclaration publique de la vice-présidente est une surprise pour les agents communautaires puisque les élus n'avaient auparavant jamais manifesté d'intérêt pour le sujet de l'*open data*. Elle marque toutefois une « commande » du politique à l'administration. L'*open data* s'ajoute aux deux premières problématisations de la mise à disposition des données publiques de la communauté urbaine de Lyon.

*

¹ Source : <http://www.rue89lyon.fr/2012/01/12/lyon-est-elle-fermee-a-louverture-des-donnees-publiques> (Consulté le 18 août 2015)

² Source : Journal de terrain – Réunion avec Clément (23 novembre 2012)

* *

Ces trois arènes de problématisation de l'ouverture des données représentent autant de scènes sur lesquelles sont racontés des récits alternatifs joués par différents acteurs. Inspire, la Centrale de mobilité et l'*open data* sont trois arènes dans lesquelles interviennent des acteurs variés (le service géomatique, les chefs de projet de la direction de la voirie, les acteurs de la société civile), dont les intérêts et les problèmes à résoudre divergent, mais qui vont pourtant mettre en avant une solution identique : une infrastructure de diffusion de données publiques de la communauté urbaine de Lyon.

Les trois problématisations de la diffusion de données publiques sont autonomes. Elles s'inscrivent dans une même temporalité, mais elles sont portées par trois coalitions de causes qui coexistent indépendamment. S'ils appartiennent tous deux à la communauté urbaine de Lyon, le service Igéo de la direction des systèmes d'information et le service mobilité urbaine de la direction de la voirie n'ont aucun lien entre eux. Le projet de Centrale de mobilité fait référence au mouvement de l'*open data* qui émerge parallèlement au sein d'autres institutions, mais ses promoteurs sont très critiques vis-à-vis de cette problématisation de l'ouverture des données jugée idéologique et naïve¹. Enfin, les projets de diffusion des données géographiques imposés par la directive INSPIRE ne s'inscrivent pas dans les projets *open data* malgré la proximité de leurs réflexions². Cette triple problématisation s'explique par la diversité des « propriétaires³ » du problème. L'ouverture des données est le résultat d'initiatives sectorielles autonomes. Les géomaticiens du service d'information géographique d'un côté, les chefs de projet de la direction de la voirie de l'autre, et les acteurs de la société civile qui trouvent de faibles relais au sein de la direction informatique de la communauté urbaine.

¹ Comme en témoigne cet extrait d'entretien avec le responsable du projet (8 janvier 2013) : « *Après, on refuse le dogme selon lequel les données devraient être mises à disposition gratuitement. J'ai rencontré Libertic, ils sont dans une démarche très idéologique, en mode « il faut libérer les données qui sont prisonnières ». On se croirait en mai 68 ! Et ils ne se rendent pas compte qu'ils font le travail de lobby au niveau local, que font les grands groupes à Bruxelles ».*

² Dans ces collectivités, les démarches d'*open data* sont également autonomes de la publication des données géographiques imposée par la directive INSPIRE. Très souvent, il s'agit de démarches indépendantes avec la constitution de deux infrastructures de diffusion ne comprenant pas les mêmes données. Cela est notamment le fait d'une séparation organisationnelle : les initiatives INSPIRE sont le fait des services géomatiques, tandis que les démarches *open data* sont souvent portées par les services informatiques, numériques ou de développement économique.

³ J. Gusfield, *La culture des problèmes publics*, op. cit., p. 11.

La pluralité des mondes sociaux et la diversité des acteurs engagés produisent des cadres d'interprétation divers du partage des données publiques. Des récits alternatifs du problème à résoudre par la diffusion des données publiques cohabitent sans qu'aucun n'ait a priori la primauté sur les autres. La mise à disposition des données doit tour à tour résoudre le problème de l'absence de représentation partagée de l'espace communautaire, favoriser le report modal de l'automobile vers les transports en commun, encourager le développement de services numériques, et développer la transparence et la participation à l'action publique. Selon ces récits, les données changent de nature. Si les données peuvent être les mêmes, on ne leur assigne pas la même signification et on leur attribue des propriétés différentes.

- La donnée est d'abord perçue comme un instrument de représentation de l'espace au service des politiques publiques.
- Elle devient ensuite un vecteur d'information pour les usagers en mobilité de l'agglomération.
- Puis, on lui assigne un caractère économique : la donnée est alors le carburant d'une nouvelle économie numérique.
- Enfin, la donnée devient un vecteur de transparence et un instrument permettant la participation des citoyens à l'action publique.

Si les problèmes sont hétérogènes, deux points convergent dans ces différents récits. En premier lieu, le Grand Lyon est catégorisé comme l'acteur public légitime pour prendre en charge ce problème. Aucune institution publique ne semble constituer une alternative crédible pour le résoudre. La région Rhône-Alpes, le département du Rhône ou les communes du Grand Lyon ne se positionnent pas sur des projets similaires d'ouverture de données publiques¹.

En second lieu, la mise en place d'une infrastructure de diffusion des données par la communauté urbaine de Lyon apparaît comme une solution commune à l'ensemble des problèmes évoqués. Elle peut certes prendre plusieurs noms selon les récits (plateforme,

¹ Ces collectivités ne sont d'ailleurs pas identifiées comme des acteurs pouvant mettre en place une politique d'*open data*. Contrairement à la communauté urbaine de Lyon, les membres du collectif OpenData69 ne cherchent pas à les rencontrer. Un article publié en janvier 2012 par le site internet Rue89 Lyon faisant le point sur l'état de l'ouverture des données au sein de la métropole lyonnaise ne mentionne pas davantage ces collectivités.

<http://www.rue89lyon.fr/2012/01/12/lyon-est-elle-fermee-a-louverture-des-donnees-publiques> (Consulté le 18 août 2015)

entrepôt, portail, centrale de mobilité, etc.), mais le principe est similaire : une infrastructure technique permettant de rassembler les données des systèmes d'information communautaires et de les diffuser à des acteurs externes à l'institution.

Face à cette conjonction de projets, la communauté urbaine de Lyon va mettre en place un programme commun d'action publique autour de la mise à disposition des données en « *fusionnant les différents projets de création de plateformes pour défendre une vision générale, avec une seule plateforme de diffusion de données pour le Grand Lyon* »¹.

Il y a eu plusieurs éléments qui ont conduit à produire un projet groupé, à savoir le contexte open data national, plusieurs initiatives en interne qui visaient à ouvrir les données dans le cadre de projets : Optimod et la démarche Inspire. Tous ces éléments ont fait que la DSIT a envisagé de faire un projet qui fédérait ces différentes initiatives. [...] On a essayé de structurer ces différentes choses, et ce qu'il en est ressorti, c'est que plutôt que d'avoir des plateformes spécialisées, qui répondaient à un besoin géomatique, un besoin d'open data, un besoin de diffusion de données, on a envisagé une approche globale, en disant la plateforme que l'on va faire va servir à faire circuler les données dans la ville intelligente, va diffuser les données à l'extérieur, que ce soit des citoyens, des entreprises, va traiter l'infrastructure de données géographiques, va traiter les données alphanumériques, les données géomatiques, tout type de sujets. Il y avait vraiment cette approche une réponse unique pour tous les besoins².

Ainsi, si la question de l'ouverture des données est initialement problématisée dans des arènes autonomes, par des groupes d'acteurs aux intérêts divergents, elle va être envisagée de manière commune par la création d'un projet spécifique visant à apporter une « réponse unique » pour tous ces besoins. Une équipe projet, constitué de deux membres indépendants des arènes initiales va être constituée pour mettre en place cette infrastructure commune. Elle va alors entamer un nouveau travail de problématisation de la politique d'ouverture des données afin d'aligner les différents enjeux, acteurs, intérêts et instruments autour de ce projet commun.

¹ Source : Journal de terrain – Entretien avec Nicolas, direction des services informatiques, communauté urbaine de Lyon (16 janvier 2013)

² Source : Entretien avec François, chef de projet, communauté urbaine de Lyon (10 février 2015)

II. Le projet « ouverture des données publiques du Grand Lyon » : composer avec des intérêts hétérogènes

La première partie de ce chapitre a mis en avant trois arènes de formulation des problèmes que doit résoudre la diffusion des données publiques, laissant entendre qu'à une pluralité de problèmes était associée une seule et même solution. Pourtant, la mise à disposition des données publiques n'est pas une solution aussi univoque qu'elle semble l'être. Elle peut en effet prendre des formes multiples en fonction des objectifs qu'on lui assigne. Comme l'indique Philippe Zittoun, il faut cesser de considérer que les solutions sont « *des instruments neutres et stables qui n'appellent aucun travail définitionnel spécifique* »¹. Les solutions ne sont pas plus neutres que les problèmes publics auxquels elles doivent répondre. Le suivi du travail définitionnel réalisé par les acteurs en charge de la plateforme de diffusion de données de la communauté urbaine de Lyon souligne les incertitudes et le bricolage nécessaire à l'association de ces multiples problèmes à cet instrument unique.

Pour décloisonner ces arènes, un projet transversal de diffusion de données est mis en place en janvier 2012. Une équipe projet, composée d'un chef de projet technique, appartenant à la direction des systèmes d'information, François, et d'un chef de projet « usagers » appartenant à un service de la délégation aux ressources chargé des projets numériques, Clément, est mise en place afin de construire ce projet unique d'« ouverture des données du Grand Lyon ». Au sein du Grand Lyon, les trois arènes initiales deviennent alors une seule et même arène.

Ces chefs de projet vont jouer un rôle de médiateur entre les différents groupes d'acteurs. La création d'un projet unique de mise à disposition des données de la communauté urbaine de Lyon a conduit à dessaisir les « propriétaires » des projets préexistants. Il est alors nécessaire de parvenir à les intéresser afin de renforcer leur association à ce projet unique. En effet, ce lien ne sera maintenu que si la « solution » proposée correspond effectivement à leurs besoins. Or, les besoins sont multiples et hétérogènes. L'équipe projet va devoir effectuer un

¹ P. Zittoun, *La fabrique politique des politiques publiques*, op. cit., p. 75.

travail délicat de composition¹ de ces intérêts hétérogènes afin de les faire tenir au sein d'une politique unique de diffusion des données publiques de la communauté urbaine de Lyon. D'instrument au service de multiples politiques, la diffusion des données va devenir une politique publique autonome.

Pour suivre ce travail de problématisation de la politique d'ouverture des données de la communauté urbaine, j'ai choisi de m'arrêter sur trois moments d'épreuves qui traduisent la politique, les données et les acteurs intéressés. Une première phase exploratoire voit les chefs de projet interroger les acteurs potentiellement intéressés par le sujet de l'ouverture des données. Puis, la problématisation de la politique est traduite pour correspondre aux objectifs des démarches de ville intelligente et de la Centrale de mobilité. Enfin, une fois cette stratégie définie, elle est présentée aux élus de la communauté urbaine pour validation. A l'issue de ces trois moments, la politique de diffusion des données mises en place par la communauté urbaine s'éloigne des principes de l'open data et des politiques mises en place au sein des autres collectivités françaises.

A. Explorer les possibles de l'ouverture des données

Au début du projet, l'univers des possibles est encore très ouvert. On peut ouvrir les données pour favoriser la transparence, l'image du Grand Lyon, la participation citoyenne, les nouveaux services, le développement économique, se conformer à la directive INSPIRE, répondre aux besoins des communes, centraliser les données mobilité. Les premiers documents présentant le projet d'ouverture des données font référence à ces multiples possibles.

¹ Le concept de composition, proposé par Bruno Latour, traduit le travail nécessaire d'association d'intérêt, de valeurs, d'approches et d'entités hétérogènes pour obtenir un monde commun. Le résultat de ce travail de construction, incertain par nature (une composition peut échouer), est toujours fragile et révisable. La question centrale n'est alors pas de savoir si quelque chose est construit ou non, mais s'il est bien ou mal construit. Cf. Bruno Latour, « An Attempt at a "Compositionist Manifesto" », *New Literary History*, 2010, vol. 41, n° 3, p. 471-490.

Les enjeux du projet sont les suivants :

- **Développement économique** avec la participation de sociétés tierces qui peuvent créer des services à valeur ajoutée ou développer leurs offres actuelles (sur les données fournies par le GL).
- Création de **nouveaux services** pour les usagers/citoyens
- **Transparence** du GL vis-à-vis des citoyens
- **Image** de marque du GL
- Respect des **politiques publiques** (attention au mauvais usage des données)
- Meilleure **évaluation** des politiques publiques et donc meilleures prises de décision
- **Simplification** des opérations de **diffusion** des données pour les services
- Garantir aux services du GL de continuer à disposer de données externes acquises par l'intermédiaire de conventions.
- Le respect du contexte réglementaire (CADA, Inspire...)
- a plus long terme, coproduction de données ?

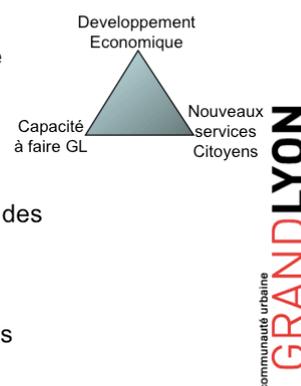


Illustration 15 : Document de présentation du projet open data du Grand Lyon

A ces enjeux multiples, sont associés autant d'acteurs hétérogènes. La première phase exploratoire engagée par les chefs du projet va consister à sélectionner les acteurs jugés légitimes pour participer à l'élaboration de la politique publique. Leurs rencontres et échanges vont délimiter les limites du « forum constituant¹ » de la politique d'ouverture des données. Du côté des producteurs de données, les directions de la prospective, de la voirie, des systèmes d'information et du développement économique sont associées, tandis que les directions de la propreté, de l'eau et du développement urbain ne participent pas à la problématisation de la politique d'ouverture des données. Du côté des réutilisateurs externes, les communes membres de la communauté urbaine et des entreprises « partenaires » vont participer à ce forum constituant. A contrario, contrairement à d'autres collectivités françaises², le mouvement citoyen local en faveur de l'open data ne contribuera pas à la politique publique d'ouverture des données.

¹ A la suite de T. Pinch et H. Collins, Michel Callon distingue deux forums dans lesquels s'élaborent les controverses scientifiques : le forum constituant dans lequel se trouvent « les quelques acteurs qui déterminent les problèmes, les arguments et les intérêts légitimes », et le forum officieux, qui rassemble les acteurs non invités dans le premier.

M. Callon, « Pour une sociologie des controverses technologiques », art cit.

² Samuel Azoulay, *De l'open data à l'open government. Le SGMAP et Etalab dans la politique de modernisation de l'action publique.*, Université de Paris II, Paris, 2014, p. 27.

1) Rencontrer les producteurs de données

Pour cadrer le projet, les deux chefs de projet organisent une série de réunions au cours desquelles des agents sont invités à répondre à la question du pourquoi de l'ouverture des données. Ces réunions regroupent une petite vingtaine d'acteurs de diverses directions (Développement économique, développement urbain, systèmes d'information, directions de l'eau, de la voirie et de la propreté, communication, prospective et dialogue public etc.). Les chefs de projet leur demandent d'« identifier les différents usages pressentis en donnant les éventuelles contraintes, et si possible classer les usages selon trois axes : développement économique, transparence, image ». L'objectif est d'« identifier les grands usages principaux pour ne pas faire fausse route et pouvoir monter une stratégie réaliste. Le recueil actuel ne doit pas limiter les usages ultérieurs ».

La deuxième séance de travail débute par un constat général : « le partage interne des données apporterait déjà beaucoup en terme de travail des services ». Ensuite, chaque représentant des directions présente les usages et objectifs qu'il a identifiés. Ceux-ci sont très hétérogènes : image de l'institution, transparence, participation, recherche universitaire, évaluation, services aux citoyens, développement économique, etc. Chacun des participants met également en avant des risques potentiels à prendre en compte lors de l'ouverture des données : qualité des données, relations avec les partenaires du Grand Lyon, appropriation des données par des « lobbys » et création de monopoles, incertitude sur l'utilisation des données, risque de mauvais usage, capacité des entreprises à exploiter commercialement les données du Grand Lyon, etc.

Ces premières séances de travail écartent durablement certains acteurs de la politique d'ouverture des données. Les directions de l'eau, de la propreté et de l'urbanisme ne perçoivent que peu d'intérêt à la diffusion de leurs données en dehors des obligations INSPIRE. « Aujourd'hui, la DGDU ne comprend pas l'intérêt de SmartData, ou alors ils le comprennent mais ils n'ont aucune envie d'y participer. Pour eux, il s'agit d'un projet Inspire, d'ouverture des données géographiques, donc, pas de leur compétence¹. » Les membres de la délégation générale au développement urbain conçoivent en effet les données avant tout comme des instruments au service des politiques d'aménagement du territoire. Ils les utilisent dans le cadre de modélisations ou d'expertises mais ne manifestent pas d'intérêt à

¹ Source : Journal de terrain – Réunion avec Clément (23 novembre 2012)

la création de nouveaux services ni au développement économique. Ils sont au contraire plutôt méfiants vis-à-vis des acteurs privés auxquels ils attribuent des volontés de captation et de « privatisation » de leurs données¹. Plutôt que d'ouvrir les données, leur crainte est que ces entreprises se retrouvent dans une situation de monopole de la production de connaissances sur le territoire au détriment des capacités d'expertise de la collectivité².

Réfractaires à toute initiative d'ouverture de données, ces acteurs ne souhaitent pas participer à la problématisation de cette politique. Leurs arguments ne sont donc plus défendus au sein du forum constituant. Les directions de la voirie et du développement économique peuvent alors imposer leur vision de la mise à disposition des données. La direction de la voirie bénéficie d'un poids important dans ce forum par la maturité de sa réflexion sur le sujet. Elle fournit en effet aux chefs de projet une note dans laquelle elle propose des « éléments de stratégie de mise à disposition des données au Grand Lyon ». Dans cette note, elle éloigne la politique d'ouverture des données de la communauté urbaine des principes de l'*open data*, qui ne sont qu'une modalité parmi d'autres de mise à disposition de données. « *D'emblée, il convient de préciser que l'Open Data est une modalité de mise à disposition des données et non une fin en soi. D'autres modalités existent (convention, licence, etc.) et la réflexion ne doit pas s'enfermer sur l'open data, le sujet étant la mise à disposition des données publiques du Grand Lyon*³. » Cette première série de réunions

¹ Ce sentiment est lié aux difficultés qu'ils rencontrent pour obtenir des données provenant des entreprises de services urbains pour établir leurs projets de modélisation urbaine.

² Cette position est partagée par certains universitaires lyonnais dont la voix ne sera pas davantage prise en compte dans la définition de la politique d'ouverture de données. « *La donnée, c'est l'or noir du 21^e siècle. Or actuellement, il y a effectivement des tentatives de privatisation de la donnée ; c'est préoccupant. C'est comme si, avec le recul de plusieurs siècles d'industrialisation, nous analysons l'impact d'avoir privatisé la matière première. Nous savons tous à quel point la possession de la matière première peut être génératrice de conflits. Et bien là, nous sommes confrontés à la privatisation de la donnée, c'est-à-dire ce sur quoi vont se construire les nouveaux services de demain. Nous ne savons pas encore exactement ce que nous allons pouvoir faire avec les données. Mais il faut préserver cette matière première car ensuite, ce ne sera pas la peine d'exécuter de nouveaux services sur des données qui n'existeront pas ou pour lesquelles il faudra payer cher pour y accéder. Une nouvelle économie pourrait se développer mais pas forcément dans le sens de l'intérêt citoyen. [...] Des représentants de grands groupes déclarent publiquement que les données publiques qu'ils peuvent récupérer ne sont plus vraiment publiques parce qu'ils les structurent, les nettoient ou les enrichissent. Ils sont en train de revendiquer que les données publiques qu'ils vont aspirer peuvent leur appartenir. Il est étonnant de voir à quel point les propriétaires de données, ceux qui collectent les données, mésestiment les possibilités d'exploitation qui pourront en être faites.* »

Source : Interview de Atilla Baskurt et Jean-François Boulicaut, <http://www.millenaire3.com/interview/liris-le-marche-des-donnees> (30 novembre 2012). Consulté le 10 août 2015.

³ Source : Note interne « Eléments de stratégie de mise à disposition des données au Grand Lyon », Direction de la voirie (19 mars 2012)

exploratoires met en avant la pluralité des enjeux que peut recouvrir l'ouverture des données publiques. Cela se confirme lors des rencontres avec les réutilisateurs de données.

2) Rencontrer les réutilisateurs de données

Parallèlement aux réunions organisées avec les services du Grand Lyon, les chefs de projet rencontrent plusieurs réutilisateurs potentiels des données ouvertes afin d'identifier leurs besoins. Le mouvement citoyen OpenData69 est tout d'abord interrogé, mais il est rapidement marginalisé au profit d'entreprises partenaires avec lesquels le Grand Lyon maintiendra un lien étroit tout au long de la problématisation de cette politique publique.

L'association OpenData69 est associée brièvement à l'élaboration de la politique d'ouverture de données du Grand Lyon, lors de la phase initiale de cadrage de celle-ci. François rencontre ainsi, à deux reprises en mars 2012, Damien Clauzel, Samy Rabih et Marie-Cécile Paccard, membres du collectif OD69. Ces réunions visent à identifier les usages possibles et les données attendues par les membres de cette association. Ils sont alors considérés, par les agents du Grand Lyon, comme des porte-parole des réutilisateurs des données publiques. Le collectif va leur fournir un document listant des exemples de données dont ils souhaiteraient l'ouverture (déplacements, développement économique, cadre de vie, environnement, état civil, etc.). Néanmoins, au delà de ces deux réunions, cette association ne sera par la suite plus conviée par les agents du Grand Lyon à l'élaboration de la politique publique. Les quelques contacts suivants seront à titre purement informatif, dans une logique descendante du Grand Lyon vers OpenData69. L'identité de cette association est en effet remise en question par les agents du Grand Lyon. A l'issue de la phase de « problématisation » de l'ouverture des données, elle n'est plus considérée comme un « porte-parole » légitime des réutilisateurs de données, mais uniquement comme des « geeks ». En réduisant ses membres à leur passion de l'informatique, le collectif OpenData69 n'est plus identifié comme une mobilisation « citoyenne ». Ils sont alors cantonnés à un forum subalterne¹, soigneusement séparé du forum constituant où s'élabore la politique d'ouverture des données.

A contrario, d'autres « réutilisateurs » sont intégrés au forum constituant. Il s'agit essentiellement d'entreprises ou de groupement d'entreprises (cluster, pôles de compétitivité)

¹ Ils continuent de se réunir à intervalles irréguliers et échangent via une liste de diffusion des informations relatives à l'ouverture des données, notamment au sein de la région lyonnaise.

avec lesquels le Grand Lyon a des relations antérieures. Par le biais de la DGDEI, et notamment de sa direction des services aux entreprises, les membres du projet SmartData rencontrent le cluster *Edit* et l'association *La Cuisine du Web*. Dans le cadre du projet Optimod, porté par la direction de la voirie, le Grand Lyon échange avec de nombreuses entreprises utilisatrices de ses données, qu'elles soient membres du projet (Cityway, IBM, Phoenix ISI, Renault Trucks) ou non (ViaMichelin, Médiamobile). Enfin, l'équipe projet est à l'écoute des retours fournies par plusieurs entreprises « partenaires » du Grand Lyon, notamment Veolia et CosmoCompany avec lesquelles des projets de modélisation urbaine sont alors en cours.

Les réutilisateurs invités au forum constituant sont ainsi ceux dont l'identité correspond à la problématisation de l'ouverture des données du Grand Lyon. Il s'agit de quelques entreprises, considérées comme des « partenaires », qui utilisent les données du Grand Lyon, afin de mettre en place des services « innovants », et qui, donc, ne remettent pas en question les problèmes et les intérêts légitimes définis par le forum constituant. Tous les autres « réutilisateurs » des données, qui ne sont pas considérés comme « légitimes » (citoyens, universitaires, développeurs indépendants, etc.), ne seront que marginalement associés à l'élaboration de la politique d'ouverture des données du Grand Lyon.

*

* *

Au terme de cette première phase exploratoire, les chefs de projet, qui ont alors peu de temps à consacrer à ce projet¹, s'appuient sur l'expertise que leur apporte la direction de la voirie sur l'ouverture des données et la légitimité fournie par la démarche de ville intelligente de la délégation au développement économique.

¹ « Quand j'ai été nommé chef de projet open data, j'avais d'autres projets en cours, je ne pouvais pas lui consacrer beaucoup de temps, il passait après les autres. Durant les premiers mois, moi, comme François, on le faisait passer après tous nos autres projets, on laissait trainer le projet. On faisait nos points projets dans la queue de la cantine ».

Source : Journal de terrain - Réunion avec Clément, notes de terrain (23 novembre 2012).

B. Définir une stratégie : « Ce n'est pas un projet d'open data, c'est un projet d'ouverture des données »

« La réflexion de la stratégie de mise à disposition des données a été fortement poussée par le projet Optimod, en lien avec la démarche Smart City également¹. » Comme l'indique ce chef de projet², la stratégie de mise à disposition des données est influencée par le projet Optimod et la démarche Smart City auxquels elle est associée.

La démarche *Smart city* trouve son origine dans le plan d'action de l'administration intitulée « La Fabrique de la Métropole ». Cette feuille de route a été présentée aux cadres du Grand Lyon le 1^{er} juillet 2011 par le directeur général nouvellement renommé Benoit Quignon. Il souhaite donner une nouvelle impulsion à l'institution en l'ancrant dans une perspective métropolitaine. Un des piliers de ce plan d'action est la « ville intelligente et durable »³. Celle-ci est définie comme le résultat de plusieurs évolutions « *technologiques, économiques, institutionnelles, environnementales, sociétales...* [qui] *viennent plus que jamais interroger, perturber, stimuler, remettre en perspective nos activités*⁴. » Une réflexion stratégique est préconisée pour positionner le Grand Lyon face à tous ces défis de la « smart city » entendue comme « *l'apport de technologies innovantes pour apporter des réponses novatrices et durables aux différents enjeux*⁵. » Cette démarche transverse à l'administration est pilotée par la délégation générale au développement économique, qui met en place un cadre stratégique et coordonne les différents projets entendus dans cette définition. Si elle est portée par la direction des services informatiques, la diffusion des données publiques est l'un de ces projets, ce qui lui vaut un rattachement décisionnel au Copil « SmartCity et services numériques ».

Le projet Optimod et la démarche ville intelligente contribuent à préciser la problématisation de l'ouverture des données autour de plusieurs points : le développement

¹ Source : Entretien avec François, chef de projet, communauté urbaine de Lyon (11 février 2015)

² Le poids de ces deux projets se manifeste notamment lors des comités de pilotage de ce projet. Si les membres de la direction des services informatiques sont les plus nombreux lors de ces réunions, les prises de parole et les décisions sont principalement le fait du directeur de la DSIT et des représentants des projets Optimod et ville intelligente.

³ Cette feuille de route, nommée « La fabrique de la métropole », contient quatre piliers interdépendants : l'ambition métropolitaine, la ville intelligente et durable, outils de l'agilité, et acteurs et artisans.

⁴ Présentation du pilier « Ville intelligente et durable » par Mathis lors de ce séminaire des cadres (1^{er} juillet 2011)

⁵ Ibid.

économique, la mise en place d'une plateforme à l'échelle de l'agglomération, l'*open data* comme élément de marketing territorial et la volonté de conserver une maîtrise sur l'utilisation des données ouvertes. La problématisation de la politique de diffusion de données de la communauté urbaine de Lyon s'éloigne alors des principes normatifs de l'*open data* pour privilégier des modalités multiples de mise à disposition des données publiques. L'institution est alors présentée comme une « Autorité organisatrice » régulant les « services publics marchands » issus des données ouvertes.

1) La Smart City : nouveaux services et développement économique

L'inscription de la plateforme *open data* au cœur de la démarche de ville intelligente du Grand Lyon conduit à associer définitivement l'ouverture des données à l'innovation et la création de nouveaux services par des acteurs privés. En effet, la démarche de « ville intelligente » est avant tout un programme de politique économique qui vise à faire du Grand Lyon un territoire d'innovation et d'expérimentation pour les entreprises du secteur du numérique. Les données sont alors perçues comme une ressource essentielle au développement de l'économie numérique.

Dans cette logique de « smart city », l'ouverture des données est envisagée comme une modalité de coproduire de nouveaux services à destination des citoyens. Identifiées comme une ressource essentielle au développement des ressources numériques, la mise à disposition des données doit permettre à des acteurs externes (entreprises, universitaires, associations, etc.) de s'en emparer et de fournir des services que la collectivité n'aurait pas imaginés.

La première grande famille d'usage de l'*open data* correspond au développement économique, aux nouveaux services pour les citoyens. C'est ce qu'on pourrait appeler un triangle vertueux entre la collectivité, les entreprises et les usagers. L'idée est la suivante : nous mettons à disposition des données, à un coût marginal puisque nous avons déjà ces données. Les entreprises les utilisent, les croisent avec d'autres qu'elles peuvent avoir par ailleurs pour développer de nouveaux services – plutôt marchands. Ainsi, les usagers disposent de plus de services numériques nouveaux et certainement plus riches que si nous les avions produits nous-mêmes puisque l'*open data* devient un levier de développement. [...] L'*open data* va ouvrir de nouveaux champs de services et de nouvelles opportunités que nous n'imaginons pas du tout. Une vraie créativité peut se manifester. Si nous faisons une étude de besoin, comme sur les projets habituels, nous limiterions complètement la créativité. Il faut laisser la porte ouverte à des choses

sur lesquelles on n'a pas la maîtrise, qui vont arriver grâce à l'interaction nouvelle de tous les acteurs. C'est vraiment une posture nouvelle pour des institutions qui sont plutôt top/down¹.

La fourniture de nouveaux services aux citoyens par des acteurs externes à l'administration est l'objectif principal associé à l'ouverture des données. Cette focalisation sur les services s'insère dans un changement de paradigme des politiques publiques de la communauté urbaine. Focalisé sur la mise en place d'infrastructures techniques, le Grand Lyon souhaite désormais s'inscrire dans une culture du service aux citoyens.

Nous sommes en train de passer d'une logique technique à une culture du service. Le Grand Lyon est une belle institution qui compte beaucoup d'ingénieurs et de techniciens qui ont développé un énorme capital de compétences techniques. Mais aujourd'hui, ce qui prime dans la gestion de la ville, c'est la qualité du service, la réactivité, l'adaptabilité par rapport aux attentes. La Communauté urbaine peut avoir tendance à faire des projets magnifiques, mais qui ne correspondent pas forcément aux attentes des citoyens. [...] L'open data s'inscrit dans quelque chose de plus large : la transformation d'une activité qui est très orientée vers la maîtrise technique à une activité orientée vers le service. C'est une grosse révolution de la structure dont l'open data n'est que la conséquence².

Pour répondre à ce nouveau paradigme de la « ville servicielle », le recours à des acteurs externes, jugés plus « agiles », « réactifs » et « innovants » que l'administration, est préconisé. Dans cette perspective, l'ouverture des données est une modalité d'inscription de la collectivité dans cette culture du service.

En outre, l'ouverture des données répond également à des objectifs de développement économique dans la mesure où la mise en œuvre de ces services par des entreprises encourage l'activité économique sur le territoire. Ces enjeux de politique économique ne sont ainsi qu'une des conséquences du développement de services et non l'objectif principal de la mise à disposition des données. La réaction d'un des chefs de projet à la première rédaction d'un document de présentation de la démarche met en avant cette hiérarchisation des priorités assignés à l'*open data*.

Clément a souhaité changer le titre de la dernière partie "*Un ensemble de dispositifs pour favoriser l'utilisation commerciale de ces données*" qui est devenu : "*Un ensemble de dispositifs pour favoriser l'innovation et la création de services.*" Selon lui, « *notre objectif est aujourd'hui de favoriser les services à l'usager, pas le développement économique. Que le service soit gratuit ou payant, on s'en*

¹ Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

² Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

fout. Notre objectif, c'est les services aux citoyens, qui peuvent générer du développement économique. Pas directement le développement éco¹. »

La distinction effectuée par Clément révèle les deux chaînes de causalité qui relient l'ouverture des données au développement économique. Soit, l'ouverture des données vise à l'émergence de nouveaux services, qui peuvent éventuellement contribuer à une création de valeur économique. Soit, l'ouverture des données favorise le développement de l'économie numérique laquelle se manifeste notamment par la création de nouveaux services pour les citoyens.

Rarement explicitées, ces deux articulations des données ouvertes et du développement économique coexistent dans la politique d'ouverture des données de la communauté urbaine de Lyon. Les acteurs associent dans un même mouvement open data, développement économique et nouveaux services, sans que le lien de causalité entre ces éléments soit clairement formulé.

Prises dans un sens ou dans l'autre, ces causalités restent néanmoins largement hypothétiques. Les acteurs ne sont pas dupes de ce caractère incertain. En s'appuyant sur les retours d'expérience d'autres collectivités, ils tentent de renforcer ce lien entre données publiques, développement économique et nouveaux services en associant plus étroitement les réutilisateurs de données.

Il faut se rappeler que le principal objectif de l'ouverture des données est le développement économique. Or, c'est une théorie qui n'a pas été démontrée par les faits. Sur les mises à disposition de données réalisées [par d'autres collectivités], très peu de services ont été créés qui soient autonomes des finances publiques. Nous-mêmes donnons depuis plusieurs années nos données de mobilité à un certain nombre d'opérateurs qui ont beaucoup de mal à en vivre. Ce sont des modèles économiques fragiles².

C'est un modèle idyllique. Dans la réalité nous n'avons pas encore la démonstration d'un modèle économique viable de l'open data pour les entreprises qui vont développer des services. [...] Nous ne savons pas encore comment ces développements de services peuvent devenir rentables³.

¹ Source : Journal de terrain, Réunion avec Clément (7 juin 2013)

² Source : Interview de Jean Coldefy, coordinateur des programmes gestion de trafic et transports publics, service mobilité urbaine du Grand Lyon, <http://www.millenaire3.com/interview/services-de-mobilite-et-donnees-publiques> (1er novembre 2012), Consulté le 10 août 2015.

³ Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

La faiblesse perçue des réutilisations commerciales des données mises à disposition par d'autres collectivités incite les agents du Grand Lyon à ne pas uniquement diffuser des données, mais à associer plus étroitement leurs utilisateurs à la politique de diffusion de données.

Les données vont être accessibles via une plateforme. Si nous nous contentons de mettre les données sur cette plateforme et attendre que les uns et les autres viennent se servir, ça ne générera rien du tout, ou si peu. Par « animation », nous entendons être en lien avec les pôles et clusters des acteurs numériques, etc. pour savoir ce dont ils ont besoin et leurs possibilités. Il s'agit aussi de monter des journées collaboratives avec des challenges, par exemple pour développer des applications smartphones à partir des données disponibles, etc. Il faut piquer au vif un peu l'écosystème qui pourrait être intéressé par ces données là pour générer quelque chose¹.

L'open data ne peut pas exister sans animation et relation partenariale avec ceux qui vont l'utiliser. Nous ne sommes pas partis dans l'optique de mettre des données à disposition et laisser les gens se débrouiller avec. Nous sommes en contact avec ceux qui ont des projets et nous travaillons avec eux. [...] Il est clair que s'il n'y a pas d'animation, ça va végéter. Il faut donc prendre le problème différemment, travailler directement sur la matière même de l'open data soit dans le cadre des projets, soit dans le cadre d'un living lab pour permettre à des entreprises ou des start-up de développer du service avec les usagers. [...] Il faut travailler sur l'innovation économique, partenariale et juridique pour avancer².

Ainsi, afin de s'assurer que les données diffusées soient réutilisées, on observe une « traduction » de l'identité du public visé par l'ouverture. Le public n'est plus un ensemble de réutilisateurs indifférenciés : il s'agit d'entreprises aux exigences de rentabilité avec lesquels la collectivité doit engager un dialogue pour les aider à atteindre une situation de viabilité économique. Cette logique partenariale établit un lien fort entre les producteurs et les utilisateurs des données ouvertes. Elle se démarque des principes de l'*open data* qui réduisent le rôle de l'acteur public à la simple diffusion de ses données. La nature de la relation évolue donc. Il ne s'agit plus de laisser des acteurs privés s'emparer librement des données publiques, mais d'associer ces acteurs à la collectivité par l'intermédiaire des données.

*

* *

La politique d'ouverture des données de la communauté urbaine de Lyon répond à la stratégie « ville intelligente » de l'institution. Cette dernière vise à encourager la création de

¹ Source : Interview de Timothée David, chargé de mission Ville intelligente, communauté urbaine de Lyon, <http://www.millenaire3.com/interview/la-donnee-carburant-de-la-ville-intelligente> (1er novembre 2012), Consulté le 10 août 2015.

² Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

services et le développement de l'économie numérique sur le territoire. La mise à disposition des données doit permettre de parvenir à ces objectifs en créant de nouvelles relations partenariales avec les entreprises fournisseurs de services.

2) Le poids d'Optimod : plateforme d'agglomération et respect des politiques publiques

Outre la démarche « smart city », la stratégie de mise à disposition des données du Grand Lyon est marquée par les besoins du projet Optimod porté par la direction de la voirie. Le projet Optimod se manifeste dans la définition de la politique de diffusion de données du Grand Lyon par le biais de deux itérations : la mise en place d'une plateforme d'agglomération et le souhait de conserver la maîtrise sur les données libérées. Ces choix éloignent la politique d'ouverture des données de la communauté urbaine des principes de l'*open data*, qui ne deviennent qu'une modalité parmi d'autres de mise à disposition de données.

(a) La plateforme du Grand Lyon ou la plateforme du territoire du Grand Lyon ?

Il faut que vous la renommez cette slide pour matérialiser cette démarche partenariale. Vous avez mis « *Diffusion des données du Grand Lyon* », il faut l'appeler « *Diffusion de données du territoire du Grand Lyon* »¹.

Cette remarque d'un agent administratif, participant au comité de pilotage du projet de diffusion des données souligne le changement progressif de positionnement de cette politique. Dans un premier temps, dans la continuité du projet de diffusion des données géographiques aux communes, l'ouverture des données ne ciblait que les données publiques de la communauté urbaine de Lyon. Cependant, sa problématisation évolue vers une logique d'agglomération pour prendre en compte « l'ambition métropolitaine » de la direction du Grand Lyon et les besoins du projet de Centrale de mobilité.

Pour comprendre ce positionnement, revenons un bref instant en arrière. Lors du séminaire des cadres du Grand Lyon du 1^{er} juillet 2011, Benoit Quignon fait son retour comme directeur général de l'institution. Il annonce alors une nouvelle feuille de route pour l'administration centrée autour de « l'ambition métropolitaine » de la collectivité.

¹ Source : Journal de terrain, Copil SmartData (20 janvier 2013)

Qu'on ne s'y trompe pas, je suis de retour mais il s'agit pour moi d'une nouvelle aventure qui commence, dans un contexte inédit et avec des défis renouvelés. [...] Il nous faudra, enfin, concevoir et mettre en œuvre nos politiques publiques en repensant l'articulation des échelles territoriales et dans certains cas, les niveaux auxquels nous intervenons. Il nous faut aussi intégrer des évolutions institutionnelles et de compétences. Face à ces « nouvelles frontières », le Grand Lyon pourra être amené à se repositionner dans certains domaines sur des missions de coordination, d'impulsion (« faire avec », « faire pour », ...), parfois tel un opérateur de services¹.

Le projet d'open data va répondre à ce défi de l'ambition métropolitaine de la collectivité en se positionnant comme une plateforme d'agglomération, mutualisant les données publiques du Grand Lyon et des acteurs du territoire. Clément justifie la pertinence de la mutualisation de la plateforme :

- Pertinent dans une démarche d'ouverture des données que les collectivités, qui ont des compétences différentes, et donc des données relevant de thématiques différentes, les mettent à disposition sur un même site.
- Cohérent dans la mesure où l'ouverture des données doit permettre de créer de nouveaux services pour les habitants. Or, aujourd'hui, à l'heure de la métropole, les habitants logent à un endroit, travaillent dans une autre ville et ont des loisirs dans une troisième. Les frontières administratives actuelles n'ont donc plus de pertinence pour un projet tel que celui-ci (d'autant plus que la mobilité est un sujet important de l'open data).
- Le choix de formats standards, reconnus, facilite l'interopérabilité entre les données provenant de divers fournisseurs et évite le mille-feuille administratif entre les différents acteurs institutionnels.
- Facteur de moindre coût pour les communes. La plateforme a été pensée dans son architecture technique pour être mutualisable dès sa conception².

En outre, l'association de la politique d'ouverture des données au projet de Centrale de mobilité renforce ce positionnement métropolitain de la plateforme. En effet, l'enjeu de ce système d'information multimodal est de faire fi des découpages institutionnels en rassemblant les données des différents opérateurs de transport de l'agglomération lyonnaise. Une fois centralisées et articulées, ces données doivent être mises à disposition d'entreprises afin qu'elles développent des services d'information multimodale à destination des usagers. L'objectif de cette infrastructure est de fournir une nouvelle représentation de l'espace de la mobilité urbaine afin d'encourager le report modal des individus. Il s'inscrit dans une volonté d'articuler les territoires des différentes offres de mobilité de l'agglomération.

Le Grand Lyon articule ainsi sa politique d'ouverture de données autour de la constitution d'une « *plateforme unique du territoire qui agrège les données des différents acteurs publics et privés, afin de mutualiser les moyens et permettre des gains de temps très* »

¹ Source : Discours de Benoit Quignon, directeur général du Grand Lyon, séminaire des cadres (1^{er} juillet 2011)

² Source : Journal de terrain – Rencontre avec le journaliste de Côté Cour, Côté Jardin (19 février 2013)

*significatifs aux ré-utilisateurs*¹ ». Alors que le découpage des compétences entre administration entraîne une pluralité d'acteurs publics producteurs de données, le souhait du Grand Lyon est « *d'éviter la multiplication des plateformes sur le territoire* »². Ce choix est présenté comme un atout pour les utilisateurs de données qui auront à leur disposition l'ensemble des données publiques sur une même plateforme quelle que soit l'administration les produisant. Il est en effet perçu que la valeur potentielle de l'ouverture des données réside dans le croisement de données provenant de différents producteurs. « *Cela est d'autant plus important que l'on sait aujourd'hui que la richesse économique ne réside pas en soi dans l'open data, mais dans un marché de la donnée (enrichissement, croisement, analyse, services) regroupant des données de multiples acteurs publics et privés* »³. » Ce discours mis en avant par la communauté urbaine est partagé par certains producteurs externes, comme le souligne cette personne de l'entreprise Keolis.

Si on faisait nous même de l'open data, mais uniquement en mode TCL, on s'inscrivait dans une logique qui était trop réduite, trop minimaliste. L'intérêt de l'open data, c'est aussi de croiser les données, notamment sur des espaces géographiques qui sont partagés. Donc, ce qui a été vu c'est de dire : « c'est pertinent de fournir cette information sur les transports en commun, mais il faut que ça s'inscrive dans un cadre plus global de l'agglomération ». D'où le choix de s'appuyer sur la plateforme SmartData du Grand Lyon pour venir apporter notre contenu⁴.

Encadré 7 : Le Grand Lyon : seul acteur légitime pour constituer une plateforme territoriale ?

Au sein d'autres métropoles françaises, plusieurs collectivités se positionnent sur le sujet de l'*open data*, ce qui mène à l'existence de plusieurs portails (Montpellier, Bordeaux⁵) ou à une démarche collective (Nantes⁶). A Lyon, la communauté urbaine est le seul acteur public à investir ce champ. Certaines communes, notamment la ville de Lyon, sont associées au projet par le Grand Lyon, mais aucune ne souhaite mettre en place son propre portail de diffusion de données. La région Rhône-Alpes, dont les échanges sont limités avec la

¹ Source : Note interne « Principes de mise à disposition des données sur le territoire du Grand Lyon » (6 juin 2013)

² Source : Journal de terrain, Réunion avec Nicolas, direction des systèmes d'information (16 février 2012)

³ Source : Note interne « SmartData : Point d'avancée sur le projet & Feuille de route 2014 » (10 octobre 2013)

⁴ Source : Entretien, Benjamin, chef de projet informatique, Keolis Lyon (13 août 2014)

⁵ A Montpellier, la ville et l'agglomération disposent chacune d'un portail *open data*. A Bordeaux, la ville, la communauté urbaine et le département de la Gironde ont lancé des projets simultanés. Dans chacun des cas, les données disponibles sur un portail ne sont pas référencées sur les autres.

⁶ La ville de Nantes, Nantes Métropole, le département de Loire-Atlantique et la région des Pays de la Loire proposent une plateforme open data mutualisée.

communauté urbaine, aurait pu se positionner comme l'acteur de référence centralisant les données de l'ensemble des institutions publiques régionales¹. Elle a certes mis en place une infrastructure régionale de données géographiques, ouverte aux collectivités régionales, pour se conformer aux exigences de la directive INSPIRE. Cependant, ce projet n'est pas couplé à une initiative *open data*². Une initiative régionale de plateforme existe néanmoins, mais, portée par Rhône-Alpes Tourisme, elle est sectorielle et concerne exclusivement le tourisme.

En l'absence d'autres démarches, la communauté urbaine a le champ libre pour se positionner comme la plateforme du territoire. Pourquoi aucune autre collectivité ne s'est positionnée sur le sujet de l'*open data* ? Deux raisons principales peuvent être avancées. D'une part, les élus des collectivités ne se sont pas emparés du sujet, ne percevant pas les enjeux en terme de positionnement de leurs collectivités. D'autre part, la communauté urbaine bénéficie de l'aubaine du « premier entrant ». Elle impose sa légitimité en étant le premier acteur régional à se positionner sur le sujet en ne laissant que peu de places pour les autres collectivités. En outre, sa démarche partenariale renforce sa prédominance. En proposant aux institutions territoriales de diffuser gratuitement leurs données sur sa plateforme, elle évite à chacun de ces acteurs des coûts de développement et de mise en place d'une infrastructure de diffusion.

Le succès de cette démarche partenariale est pourtant très incertain au lancement du projet. Les chefs de projet la problématisent, mais aucun ne sait si les acteurs externes vont souhaiter s'y associer. Un travail d'intéressement sera nécessaire pour parvenir à les enrôler. Les réunions avec des producteurs variés vont se multiplier (communes, CCI, opérateurs de transport, etc.), les choix graphiques de la plateforme sont soigneusement étudiés pour ne pas trop la « typer » Grand Lyon, les coûts de diffusion sont intégralement pris en charge par la communauté urbaine, etc. Pourtant, quelques semaines avant le lancement du portail, la « plateforme partenariale » ne compte toujours aucun « partenaire ». Les données des Vélov diffusées par ailleurs par l'entreprise JC Decaux sont récupérées, et une pression est exercée auprès de la ville de Lyon afin qu'elle publie quelques jeux de données géographiques.

¹ La région Provence-Alpes-Côte d'Azur a construit un portail partenarial à l'échelle de la région qui rassemble les données d'une trentaine d'acteurs publics.

² Selon un chargé de mission développement numérique, il est difficile de lancer un portail régional car la région « ne dispose pas de données. Celles-ci sont aux mains des divers prestataires qui agissent pour son compte ». La collectivité se concentre alors sur la mise en place de dispositif de médiation et de formation à la données (Infolab, etc.).

(b) Laisser-faire ou contrôler les réutilisations ?

Selon les principes édictés par les partisans de l'open data, les acteurs publics ne doivent pas discriminer les utilisateurs des données ouvertes. Cette conception libérale laisse libre cours à de nombreux usages de ces données, sources de créativité pour les services développés, mais également d'incertitude pour les acteurs publics. En effet, contrairement aux partenariats public/privé classiques, l'entreprise ne répond plus à un cahier des charges défini par la collectivité, mais elle est chargée de le définir et de le mettre en œuvre. Ce laisser-faire soulève de nombreuses interrogations chez les agents administratifs : que faire en cas d'usage qui irait à l'encontre de l'intérêt général ? Doit-on laisser aux mains du secteur privé la détermination de l'intérêt général ? Comment s'assurer que les réutilisations soient cohérentes avec les politiques publiques ?

Sous l'influence des porteurs du projet Optimod, ces questions sont illustrées par le cas de la politique de mobilité urbaine. Cet exemple va conduire le Grand Lyon à opter pour une mise à disposition régulée de ses données.

On ne peut pas se permettre qu'une donnée relative à la mobilité soit saisie par un opérateur privé qui a vocation à faire entrer plus de voitures dans des zones où l'on souhaite réduire l'usage des véhicules¹.

Si on fournit des données –par exemple de trafic en temps réel – à des opérateurs qui, grâce à l'information du Grand Lyon, constatant qu'il y a des bouchons sur une voie, envoient le trafic sur d'autres voies (sans information trafic) le long desquelles il y a des écoles et des habitations, c'est totalement incohérent avec nos politiques publiques. Nous avons donc besoin d'avoir des garanties. Il faut dans l'exemple cité que les opérateurs s'engagent à modifier leur algorithme de calcul de navigation pour s'assurer qu'ils ne nuisent pas à l'intérêt général².

Pour prévenir des usages allant à l'encontre de l'intérêt général, la mise à disposition des données doit se faire « *selon des conditions permettant de s'assurer de la cohérence des services avec la politique publique* »³. Face aux acteurs privés, la Communauté urbaine se

¹ Source : Interview de Timothée David, chargé de mission Ville intelligente, communauté urbaine de Lyon, <http://www.millenaire3.com/interview/la-donnee-carburant-de-la-ville-intelligente> (1er novembre 2012), Consulté le 10 août 2015.

² Source : Interview de Jean Coldefy, coordinateur des programmes gestion de trafic et transports publics, service mobilité urbaine du Grand Lyon, <http://www.millenaire3.com/interview/services-de-mobilite-et-donnees-publiques> (1er novembre 2012), Consulté le 10 août 2015.

³ Pour cela, il préconise un conventionnement avec « un contrôle de la qualité et de la cohérence avec les politiques publiques. Il y a un risque avec par exemple des services qui n'incitent pas au report modal, voire le contraire, des services incitant à l'étalement urbain. »

Source : Note interne « Modalités de fourniture des données au Grand Lyon » (17 juin 2013)

présente alors comme le garant de l'intérêt général, et, dans cette perspective, souhaite réguler les utilisations des données libérées.

Le fil rouge, c'est d'affirmer que nous sommes là pour donner des orientations, défendre des politiques publiques et pas pour faire uniquement le jeu des entreprises ou satisfaire des groupes d'intérêts qui pourraient être contraires à ces politiques ». [...] Un terme commence à faire son chemin à la Communauté urbaine de Lyon, c'est la notion d'« autorité organisatrice ». Nous sommes là pour défendre le service public. Il faut retrouver un contrepoids à une activité purement libérale. Ce peut être en régulant la mise à disposition des données car ces données sont nécessaires pour la gestion de la ville¹.

Cette traduction de l'ouverture des données publiques vise à la rendre compatible avec les préoccupations civiques défendues par les agents du Grand Lyon. Elle transforme la posture et le rôle de la communauté urbaine : elle n'est plus seulement productrice de données utiles au développement de l'innovation, mais elle se positionne comme une « autorité organisatrice » garante de (sa vision de) l'intérêt général. La problématisation de la diffusion des données publiques s'éloigne alors de l'open data pour inclure des modalités diverses de diffusion.

C'est un point qui a toujours été au centre de toutes les réflexions des groupes de travail que nous avons eus sur la ville intelligente : quelle est la posture d'une collectivité garante du service public dans l'open data ? Quel est notre rôle ? Peut-on faire de l'open data et laisser, quelle que soit la donnée, tout le monde l'utiliser pour faire ce qu'il en veut ? La Direction de la voirie prend souvent l'exemple d'une ouverture des données qui servirait à faire entrer davantage de voitures en ville : c'est exactement contraire à nos politiques publiques ! Il est absolument essentiel d'avoir une ouverture raisonnée, contrôlée.

Je préfère parler d'ouverture des données publiques dont l'open data n'est qu'une modalité. L'open data est un concept, presque une idéologie ! C'est une certaine vision de l'ouverture des données publiques, mais pas la seule. Nous pouvons avoir intérêt à n'ouvrir certaines données qu'à certaines populations, de manière conventionnée et maîtrisée. Pour avoir un retour, et parce qu'on veut maîtriser la politique publique.

Sur la dimension libérale de l'open data "développeur de business", nous mettons le curseur sur l'ouverture à tous, mais contrôlée au regard de nos politiques publiques. C'est notre position actuelle. Payant ou gratuit, ce sera un peu à la carte, en fonction des données. Le fil rouge de la condition d'ouverture ce n'est pas la position libérale / libertaire, c'est plutôt d'affirmer que nous sommes là pour donner des orientations, défendre des politiques publiques et pas pour faire uniquement le jeu des entreprises ou satisfaire des groupes d'intérêts qui pourraient être contraires à ces politiques. Nous essayons d'adopter une position plutôt équilibrée².

¹ Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

² Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

La communauté urbaine de Lyon s'écarte de l'*open data* pour privilégier des modalités diverses de mise à disposition de ses données afin de conserver la maîtrise de ses politiques publiques. Ce choix est renforcé par le souhait que l'infrastructure mise en place puisse répondre à l'ensemble des projets du Grand Lyon nécessitant un échange de données entre acteurs.

3) Une plateforme au service des politiques du Grand Lyon et de la ville intelligente

La démarche d'ouverture des données répond aux besoins de ses politiques publiques, notamment de fournir des informations géographiques fiables aux acteurs agissant sur le territoire, dont les communes du Grand Lyon, et à la Centrale de mobilité qui nécessite une plateforme regroupant les données de l'ensemble des modes de transport de l'agglomération afin de permettre la mise en place de services d'information multimodale. Comme l'indique un de ses documents de présentation, la communauté urbaine de Lyon ne cherche ainsi pas à « *faire de l'open data pour faire de l'open data* ».

Plusieurs raisons nous ont poussé à ouvrir d'abord les données géographiques. D'une part, la directive européenne INSPIRE nous impose de les diffuser. D'autre part, il y a une attente très forte des 58 communes du Grand Lyon à utiliser cette plateforme pour intégrer leurs propres données et les gérer. On ne connaît pas le retour sur investissement ou la viabilité de l'*open data*, mais on s'est dit que la conjonction de l'obligation INSPIRE et de la demande des communes valait le coup d'investir dans ce champ. Quand on est capable de répondre à ces deux premières demandes, on fait de facto de l'*open data*, donc ouvrons nos données¹ !

Enfin est rentré en ligne de compte, le fait que l'on était pas précurseur en matière d'*open data*, et que l'on voulait profiter des expériences d'autres collectivités, notamment sur le fait que l'*open data* en tant que tel n'était pas forcément un succès en terme de développement économique ni même appropriation citoyenne des données. Donc, on a essayé aussi d'orienter notre travail autour de la data sur je fais circuler la donnée en interne des services, aux partenaires, que ce soit les communes ou d'autres partenaires institutionnels, et aux partenaires industriels ou aux entreprises du territoire, avec une logique, là plus ouverture des données. [...] L'ouverture, pas que l'*open data*².

Son éloignement d'une démarche *open data* s'explique également par les retours d'expérience des initiatives *open data* mises en place par les collectivités françaises pionnières et l'incertitude quant au développement économique qui en découlait. Dès lors, elle va proposer une problématisation plus large autour de l'ouverture des données entendue

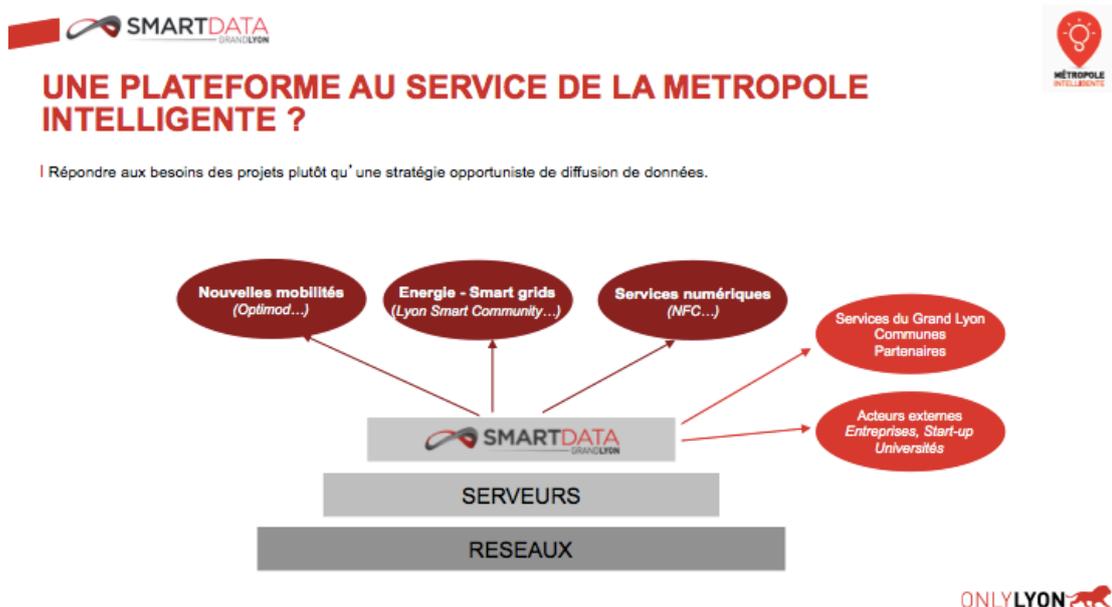
¹ Source : Interview de Philippe Vial, directeur des systèmes d'information du Grand Lyon, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (30 novembre 2012), Consulté le 10 août 2015.

² Source : Entretien avec François, chef de projet, communauté urbaine de Lyon (11 février 2015)

comme l'ensemble des modalités de circulation des données numériques entre les acteurs de l'agglomération. Ce choix se manifeste dans le nom de la plateforme « SmartData » afin de ne pas faire référence explicitement à l'open data.

La maîtrise et la circulation des données sont un enjeu crucial de la ville intelligente. En tant qu'infrastructure intermédiaire entre les producteurs et les utilisateurs de données, la plateforme SmartData peut jouer un rôle stratégique permettant au Grand Lyon de se positionner comme l'institution principale de régulation de la Métropole intelligente. A ce titre, il est crucial que l'articulation de la gouvernance de la Métropole intelligente et de SmartData soit renforcée, en positionnant Grand Lyon SmartData comme le support en matière de données pour l'ensemble des projets de la Métropole intelligente (Lyon Urban Data, Energie, etc.)¹.

L'association du projet à la démarche « Smart city » va renforcer cette perspective. En effet, afin de ne pas être marginalisé en étant perçu comme un simple projet « open data », les chefs de projet vont mettre en avant ces modalités différenciées de circulation de données pour s'inscrire comme le support de la ville intelligente. La plateforme est présentée « *comme un socle à partir desquelles viennent se greffer les différents services mis en place par le Grand Lyon ou par des acteurs privés dans les différents domaines d'intervention urbains*². » Ce positionnement de la politique d'ouverture des données au service des autres politiques de la communauté urbaine est manifeste dans le document de présentation suivant³.



¹ Source : Note interne « SmartData : Point d'avancée sur le projet & Feuille de route 2014 » (10 octobre 2013)

² Source : Document interne (Septembre 2013)

³ Source : Document de présentation de la plateforme SmartData lors du Copil SmartCity du 25 novembre 2013

*

* *

Sous l'influence des directions de la voirie et du développement économique, la communauté urbaine de Lyon traduit les principes du mouvement de l'*open data* pour les adapter à ses intérêts. La stratégie de mise à disposition définie par le Grand Lyon vise à répondre aux besoins de la démarche Ville intelligente et du projet Optimod : la diffusion des données doit encourager la création de services numériques et le développement économique du territoire, en veillant que ces réutilisations n'aillent pas à l'encontre des politiques publiques de l'institution. Dès lors, tandis que les partisans de l'*open data* préconisent de mettre à disposition les données gratuitement, librement et sans aucun contrôle de la réutilisation, le Grand Lyon souhaite mettre en place une plateforme regroupant les données de l'agglomération, contrôler l'usage de certaines données jugées stratégiques et préserver un écosystème concurrentiel équitable. Ces orientations, décidées par les équipes administratives du Grand Lyon, doivent être présentées aux élus de la communauté urbaine afin de valider cette stratégie d'ouverture des données.

C. Faire valider la stratégie : les rencontres avec les élus

En l'absence de commande politique explicite, l'administration de la communauté urbaine s'est chargée de définir les grandes orientations de la stratégie de diffusion des données sans en référer aux élus de l'institution. Le développement économique et la création de services numériques sont privilégiés. Toutefois, la transparence et la participation citoyenne restent présentes en arrière plan dans le discours des porteurs du projet. Les rencontres avec les élus du Grand Lyon vont faire évoluer les termes du débat. Le développement économique est privilégié et la fin du volet citoyen de la diffusion des données est actée.

Après quelques mois de travail, les chefs de projet rencontrent tour à tour trois vice-présidents de la communauté urbaine : Karine Dognin-Sauze, en charge des nouvelles

technologies et de l'innovation, Sandrine Frih, vice-présidente aux systèmes d'information et Gérard Claisse, chargé de la participation citoyenne¹.

Ces réunions avec les élus visent davantage à obtenir une validation des orientations déjà définies qu'à faire émerger une nouvelle problématisation du projet². Les rencontres se déroulent selon le même schéma : présentation du projet et contextualisation puis une présentation de la plateforme avec la diffusion d'une vidéo de démonstration d'une réutilisation des données dans Google Maps. Si la vice-présidente aux systèmes d'information se montre enthousiaste sur la plateforme, elle ne fait aucun commentaire sur une orientation à suivre. A contrario, Karine-Dognin Sauze manifeste davantage d'intérêt pour le sujet et soutient les orientations choisies par l'administration. Alors qu'au cours de la réunion un des chefs de projet mentionne que l'open data est un sujet à la mode, elle l'interrompt et affirme « *Le Grand Lyon ne doit pas aller dans le sens de la mode* »³. La priorité accordée au développement du tissu économique local, à l'amélioration des services rendus aux usagers et au rôle de la collectivité de régulateur des réutilisations lui convient parfaitement. Elle valide également le rattachement du projet de diffusion de données à la démarche de ville intelligente – dont elle a la responsabilité.

La diffusion des données publiques peut répondre à des enjeux de transparence et de participation citoyenne. Identifié par l'équipe en charge du projet d'ouverture des données du Grand Lyon, cet objectif est défendu par des membres de la direction de la prospective et du dialogue public. A leur initiative, une réunion est organisée avec le vice-président à la participation citoyenne afin de lui présenter le projet d'ouverture des données du Grand Lyon. Si l'élu se montre intéressé par la thématique, il ne souhaite pas s'emparer du sujet de

¹ Le projet d'ouverture des données a également été exposé au président de la communauté urbaine, Gérard Collomb, lors d'une présentation générale de la stratégie SmartCity. Selon les dires d'un des participants à cette réunion, politique à trois licences et l'objectif orienté développement économique ont bien été identifiés (et donc entérinés) par le président du Grand Lyon.

² Le rapport entre l'administration et le politique mériterait une étude dédiée. Si les élus sont peu présents explicitement dans la définition de la stratégie de diffusion de données, ils restent présents en filigrane dans la mesure où les agents anticipent leur réaction pour construire le projet. En outre, le soutien politique est recherché afin de légitimer le projet vis-à-vis d'autres directions de l'administration. Toutefois, si le pouvoir décisionnel des élus semble important, en pratique, ils sont davantage en situation de dépendance vis-à-vis de l'administration. Ils ne disposent pas ou peu de chargés de mission et sont en situation d'infériorité face à l'expertise déployée par l'administration. Cette dernière joue de ce monopole de l'expertise en présentant les projets d'une manière telle que les orientations qu'ils privilégient soient acceptées par les élus politiques. Une des techniques pour cela est par exemple l'utilisation de trois « scénarios », dont l'un est mis en avant et préconisé par les agents administratifs.

³ Source : Journal de terrain, Réunion avec Karine Dognin-Sauze (4 octobre 2012)

l'ouverture des données publiques, qui peut être sensible politiquement, à un an des élections, d'autant qu'il estime que les élus sont « *à poil là dessus, on n'a pas de vision, pas de philosophie sur la question*¹ ». Malgré l'insistance de Meryl pour que la mission participation citoyenne se positionne sur l'ouverture des données, l'élu ne souhaite pas s'engager sur le sujet de l'*open data*. Au contraire, il craint d'être interpellé sur la question dans cette période pré-électorale. Dès lors, en l'absence de porte-parole, aucune coalition ne se forme pour imposer ce volet transparence et participation citoyenne et le mettre à l'agenda de la politique d'ouverture de données.

Les rencontres avec les élus de la communauté urbaine renforcent l'orientation « développement économique » de la politique d'ouverture de données définie par l'administration. Si la problématisation générale de la politique n'est pas remise en cause, le volet transparence et participation citoyenne est définitivement abandonné.

III. Elaborer les licences de réutilisation : une mise en boîte noire de la stratégie de diffusion

La troisième partie de ce chapitre interroge les modalités de mise en boîte noire de la politique d'ouverture des données de la communauté urbaine de Lyon. Comment les problématisations ont été stabilisées ? Comment sont effacés les traces, les bricolages, les tâtonnements qui ont jalonné le processus de cadrage de la politique ?

Pour étudier les modalités de mise en boîte noire de cette stratégie, la dernière partie de ce chapitre est consacrée à la fabrique des licences qui encadrent la mise à disposition des données. Ces dispositifs réglementaires sont au cœur des politiques d'*open data* puisqu'ils cadrent les usages possibles des données ouvertes. Un des principes du mouvement de l'*open data* mentionne explicitement l'importance du choix de la licence pour caractériser une donnée d'*open*. Les militants ne s'y trompent pas : ils sont particulièrement vigilants quant aux licences choisies par les administrations n'hésitant pas à les critiquer de manière

¹ Source : Journal de terrain (15 avril 2013)

virulente¹. Si les licences sont l'objet d'autant d'attention, c'est qu'elles sont dotées d'une *agency*, une capacité d'action sur les choses et les êtres contribuant à les transformer². Les licences ne formalisent pas une politique qui serait « déjà-là », elles produisent des effets, elles s'insèrent dans la politique qu'elles contribuent ainsi à façonner, et, surtout, elles influent sur les actions futures.

Prendre cet instrument d'action publique comme point d'entrée dans l'analyse permet de retracer l'évolution puis la stabilisation de la problématisation de l'ouverture des données. Le choix et la rédaction des licences constituent une épreuve de traduction supplémentaire qui contribue à formaliser la politique d'ouverture des données et à verrouiller l'association de certains acteurs à celle-ci. Tout l'enjeu est de réussir à faire tenir ensemble les différentes significations associées à la politique d'*open data* : il faut parvenir à trouver un compromis conciliant les intérêts de l'ensemble des acteurs. Ce processus d'instrumentation de l'action publique, entendue comme « *l'ensemble des problèmes posés par le choix et l'usage des instruments (des techniques, des moyens d'opérer, des dispositifs) qui permettent de matérialiser et d'opérationnaliser l'action gouvernementale* »³, contribue à la solidification des contours des problèmes publics auxquels doit répondre l'ouverture des données.

Cette instrumentation prend la forme de trois licences, votées par le conseil communautaire du 26 septembre 2013⁴, qui incarnent les grandes orientations de la politique d'ouverture de données du Grand Lyon. La première, la « licence ouverte », en se contentant de rappeler les obligations légales⁵, répond à la volonté de favoriser le développement économique par la mise à disposition des données. La seconde, la « licence engagée », en

¹ Voir par exemple ces billets de blog des associations Regards citoyens et Libertic : *L'OpenData, c'est bien plus que de la communication politique!*, <http://www.regardscitoyens.org/lopendata-cest-bien-plus-que-de-la-communication-politique/>, 9 juin 2010, (consulté le 10 août 2015) ; *Licences OpenData : L'APIE grille la priorité à ÉtaLab et invente le pseudo-libre*, <http://www.regardscitoyens.org/licences-opendata-lapie-grille-la-priorite-a-etalab-et-invente-le-pseudo-libre/>, 19 juillet 2010, (consulté le 10 août 2010) ; *Pourquoi n'y a-t-il pas de consensus sur une licence Open Data en France ?*, <https://libertic.wordpress.com/2011/07/05/pourquoi-ny-a-t-il-pas-de-consensus-sur-une-licence-open-data-en-france/>, 5 juillet 2011, (consulté le 10 août 2010).

² Je m'inscris ici dans la continuité des travaux développés en sociologie des sciences sur le rôle des artefacts sociotechniques, en pointant leurs caractéristiques politiques.

³ Pierre Lascombes et Patrick Le Galès, *Sociologie de l'action publique*, Paris, Armand Colin, 2007, p. 104.

⁴ Source : Délibération n°2013-4095 du conseil de la communauté urbaine de Lyon

⁵ La mise à disposition des données publiques est encadrée par la loi n°78-753 du 17 juillet 1978, qui indique que les autorités publiques peuvent fixer librement les modalités de réutilisation de leurs données. Les utilisateurs de ces données sont contraints d'indiquer l'origine des données diffusées ainsi que la date de la dernière mise à jour. Ils doivent également veiller à ne pas altérer les données, ni dénaturer leur sens.

instaurant une authentification de l'utilisateur, révèle le souci de contrôler la conformité des usages aux politiques publiques. La troisième, « la licence associée », en instaurant un système dissuasif de redevances, manifeste le désir de maintenir un marché non-monopolistique. Enfin, le caractère générique de ces licences traduit la volonté d'établir une politique s'appliquant à l'ensemble des producteurs et des données de l'agglomération. Ces licences éloignent la stratégie du Grand Lyon des principes de l'*open data* qui préconisent de mettre à disposition les données gratuitement, librement et sans aucun contrôle de la réutilisation.

L'absence de débat lors du conseil communautaire masque le travail sinueux et incertain de composition d'intérêts, de valeurs et d'entités hétérogènes qui assure la robustesse et la solidité de ces licences. Le récit proposé dans cette dernière partie s'immisce à l'intérieur des coulisses de la fabrique des licences. Dans les détails de la rédaction se logent de longs fils de conséquences politiques, économiques, techniques ou organisationnelles. L'observation des points de débats et des problèmes pratiques rencontrés révèle un ensemble de discrètes décisions qui, sédimentées, entérinent la formalisation de la politique d'ouverture de données de la communauté urbaine de Lyon.

A. Associer des acteurs et des données hétérogènes en adoptant des licences génériques

La première traduction de la stratégie de diffusion des données se manifeste par la définition de licences génériques. Cette volonté de construire un cadre juridique commun traduit le souhait de la communauté urbaine de mettre en place une plateforme unique de diffusion de données à l'échelle de l'agglomération.

Aujourd'hui, l'intérêt du Grand Lyon est de proposer un cadre juridique global, s'appliquant à l'ensemble des cas d'ouverture des données publiques. Il s'agit donc de proposer un type de licence global, de ne pas multiplier les licences pour chaque direction, mais proposer un modèle commun. Les partenaires du Grand Lyon (SYTRAL, ville de Lyon) doivent proposer des licences similaires à celles du Grand Lyon afin d'en faciliter la lisibilité¹.

Auparavant chacune des directions du Grand Lyon était maître de définir sa propre licence encadrant la mise à disposition de ses données. Une grande hétérogénéité prévalait alors au détriment de la cohérence et de la lisibilité de la politique publique. Cette volonté de standardisation contribue à sortir les débats sur les licences des périmètres sectoriels

¹ Source : Journal de terrain – Réunion « Licences » (17 décembre 2012)

(géographique, mobilité) dans lesquels ils étaient cantonnés pour adopter une arène transverse. Celle-ci va regrouper les responsables des projets Optimod et SmartData, et les services juridiques du Grand Lyon et du Sytral¹. Propriétaire des données des transports en commun lyonnais, ce dernier est identifié comme un acteur indispensable à la réussite de la plateforme d'agglomération. En associant le Sytral à ces discussions préliminaires, le Grand Lyon s'assure de sa participation à son projet de diffusion de données.

L'abandon des licences sectorielles encadrant la diffusion des données conduit à dissocier le cadre juridique du type de données mises à disposition. Les licences ne sont plus spécifiques à une catégorie juridique de données : données publiques, données privées, données en temps réel, données statiques, données géographiques, données mobilité, etc. Une catégorie générique de « Données » est créée regroupant « *l'ensemble des informations publiques et des autres données, y compris leurs mises à jour, fichiers, base de données et autres informations mis à la disposition du Licencié par la communauté urbaine de Lyon* »².

B. ODbL versus Licence ouverte : une première licence pour favoriser les réutilisations commerciales

Le premier objectif assigné à la diffusion des données est d'encourager la création de services numériques et de favoriser le développement économique. Alors que le Grand Lyon avait initialement opté pour une licence ODbL, la communauté urbaine va revenir sur ce choix et adopter une « licence ouverte » jugée plus en phase avec cette orientation économique.

D'abord effectué pour répondre aux besoins de diffusion des données géographiques dans le cadre d'Inspire, le choix de la licence s'est porté initialement sur la licence ODbL. Après un rapide benchmark³, le service de l'information géographique préconise la licence ODbL qui lui « semble être celle à choisir⁴ » face aux limites des autres licences⁵. Le choix de

¹ A noter que la ville de Lyon n'est pas associée à ces discussions alors que dans le même temps on tente de l'associer au projet d'ouverture de données.

² Source : Licence Engagée de la Métropole de Lyon – Article 2

³ Source : Note interne du service Igéo « Licences applicables aux données diffusées par Igéo » (6 septembre 2011)

⁴ Source : Compte-rendu de réunion (06 mars 2012)

⁵ La licence Créative Commons est « plutôt pour les documents », la licence ouverte est « franco-française et risque de poser des problèmes pour l'extérieur du territoire » tandis que l'absence de licence n'assurerait pas une sécurité juridique suffisante aux réutilisateurs.

cette licence ODBL est validé par le vote d'une délibération par le Conseil de Communauté approuvant l'ouverture des données publiques et décidant « *la gratuité des modalités de diffusion des données publiques de la Communauté urbaine sous licence de type ODbL¹* ».

Les enjeux politiques de cette licence n'auront cependant jamais été réellement soulevés. Or, la particularité de la licence ODbL est sa clause de type « share-alike » qui oblige les utilisateurs à placer toutes bases de données dérivées sous la même licence.

François : Pourquoi ODBL ? Il y a un choix un peu opportuniste qui a été fait il y a un an parce qu'il fallait qu'on lance la plateforme et que l'on a choisi cette licence après comparaison de quelques autres licences et que ça nous semblait la bonne solution, Paris l'avait fait... mais ça n'a pas été beaucoup plus instruit et argumenté que ça. Il nous a semblé que c'était la licence qui convenait le mieux à notre projet, notamment du fait de la compatibilité avec l'international.

Enquêteur : Et quid du Share-Alike ?

François : Ce n'est pas un point qui nous a marqué, il n'est pas entré en compte dans le choix de la licence².

La clause de partage à l'identique de cette licence libre soulève des enjeux politiques qui n'ont pas été pris en considération. Placées sous cette licence, les données ouvertes deviennent des biens communs à l'ensemble de la communauté des utilisateurs. Elles ne peuvent alors être appropriées par un acteur privé pour son usage exclusif.

Parallèlement à ce choix initial de licence, la direction de la voirie commande à un cabinet d'avocats parisien, spécialisé dans l'économie numérique, une étude sur les licences à mettre en œuvre dans le cadre de son projet de Centrale de mobilité. Ce cabinet met en garde la communauté urbaine de Lyon contre la licence ODBL et préconise l'utilisation de licences spécifiques. Selon son analyse, cette licence « virale » serait incompatible avec la législation française parce qu'elle impose des contraintes à la réutilisation non prévues par la loi de 1978 et que sa rédaction en langue anglaise contredit la loi Toubon³. Surtout, la licence ODbL n'est pas perçue comme optimale pour assurer la réutilisation commerciale des données. Cette licence, initialement choisie par le Grand Lyon, est alors remise en cause par la direction de la voirie au motif qu'elle ne permettrait pas de « faire du business⁴ » avec les données ouvertes.

¹ Source : Délibération n°2012-3081 (25 juin 2012)

² Source : Réunion avec François (09 juillet 2013)

³ La loi 94-665 dite Toubon impose que les contrats conclus par les administrations soient rédigés en langue française.

⁴ Source : Journal de terrain (17 décembre 2012)

Cette analyse et l'absence de véritable débat sur les enjeux politiques des licences m'intriguent comme je le relate dans mon journal de terrain. Je vais alors tenter de susciter ce débat en produisant une note d'analyse comparant les deux licences les plus utilisés dans les projet d'ouverture de données : la licence ODbL et la « licence ouverte¹ ».

Cette réunion m'a intrigué, notamment sur le fait de l'impossibilité déclarée des réutilisations commerciales dans le cadre de la licence ODbL, et également, sur la sensation que le choix de la licence ne faisait pas réellement débat entre les différents acteurs. Je ne sais pas comment le choix initial de la licence ODbL a été effectué, ni quelles sont les raisons à l'origine de ce choix, mais, l'abandon de ce type de licence ne semble pas poser problème aux acteurs. Ils semblent se référer au travail effectué par le cabinet « d'experts » [François], sans regard critique sur celui-ci. Or, le choix d'une licence est éminemment politique : il détermine le statut de la donnée, les relations du Grand Lyon avec ses partenaires et la nature de la communauté des réutilisateurs. Il ne s'agit donc pas d'une décision à prendre à la légère.

Le lendemain, lors du point projet hebdomadaire SmartData, je fais part à Clément et François de ma surprise quant à l'absence de débat relatif au choix des licences, en leur présentant les enjeux que représente ce choix sur le type de partenariat que le Grand Lyon veut établir avec les utilisateurs. Ils m'ont dit qu'effectivement, il s'agissait d'une question politique importante, qu'ils n'avaient pas perçue, mais que cela méritait débat et validation par le politique. Clément me demande de rédiger une fiche de présentation de ces différents enjeux, afin d'obtenir un accord politique et une validation au Copil de janvier, suite à une discussion avec le service juridique et les autres interlocuteurs².

Je rédige alors une note pointant les enjeux politiques des licences qui, selon mon analyse, « *détermine[nt] l'orientation stratégique de l'ouverture des données et la nature des rapports entre la collectivité et les acteurs privés* ». Je rappelle ensuite que la licence ODbL n'empêche pas les réutilisations commerciales, mais déplace la valorisation de la donnée vers le service. Dans ce cadre, la donnée est envisagée comme un bien commun non appropriable. A contrario, la licence ouverte est présentée comme une licence « libérale » qui considère la donnée comme une ressource qui peut être valorisée pour elle-même par les acteurs privés. Par le biais de cette note, j'essaie de faire émerger une réflexion sur le positionnement du Grand Lyon vis-à-vis des acteurs économiques : doit-il « laisser faire » le marché des données ouvertes ? Ou doit-il le réguler afin d'empêcher toute privatisation des données publiques ?

La clause de partage à l'identique ne constitue en aucun cas une impossibilité pour l'utilisateur d'utiliser la donnée dans un cadre commercial. Si elle empêche la commercialisation des bases de données enrichies, elle autorise la commercialisation de services produits à partir de ces bases de données. Elle favorise ainsi l'émergence de modèles économiques basés, non pas sur la vente de droit d'accès à une base de données, mais sur la création de services à partir de ces bases de données. La licence ODbL permet ainsi à la collectivité d'orienter les réutilisations commerciales

¹ La licence ouverte est une licence proposée par la mission gouvernementale Etalab qui se contente de rappeler les obligations indiquées dans la loi de 1978.

² Source : Journal de terrain (18 décembre 2012)

de ses données vers la création de services. La valorisation est déplacée de la donnée vers le service.

L'obligation de partage à l'identique contraint les utilisateurs de la donnée à mettre à disposition d'autres utilisateurs le jeu de données enrichi. Les données sont ainsi placées dans un « pot commun », qui s'auto-entretient, dans lequel chacun des acteurs peut venir puiser pour créer de nouveaux services. L'objectif de ce type de licences « virales » est de créer un écosystème d'innovations dynamique et collaboratif, de permettre un enrichissement global de la collectivité et d'éviter l'appropriation de ces données par certains acteurs au détriment des autres. En effet, un des dangers de l'ouverture des données est de conduire à une « privatisation » des données publiques, financées par l'investissement de la collectivité, dont certains acteurs privés s'approprieraient sans fournir de contreparties à la société. Le risque que la collectivité perde le contrôle sur les données urbaines est réel. Or, la donnée est au cœur de la régulation de la future Smart City. Si le Grand Lyon souhaite être à même de préserver sa capacité de négociation avec les acteurs privés lors de la détermination des projets de la ville, il doit conserver un pouvoir et une maîtrise sur les données ouvertes.

Le choix de la licence est ainsi éminemment politique puisqu'elle détermine l'orientation stratégique de l'ouverture des données et la nature des rapports entre la collectivité et les acteurs privés. La clause de partage à l'identique correspond ainsi à une approche moins « libérale » de l'ouverture des données que celle qui est privilégiée par la « licence ouverte ». Cette dernière laisse une grande liberté aux réutilisateurs dans les modalités de valorisation des données publiques, tandis que la licence ODbL permet à la collectivité de contrôler davantage l'usage qui est fait de ses données, de préserver une équité entre les différents acteurs les utilisant, et de garder ainsi une maîtrise plus importante sur la détermination des politiques publiques. Il faut néanmoins noter que la licence de type ODbL présente l'écueil d'obliger les utilisateurs à ouvrir et partager leur propre base de données. Cet aspect peut légitimement provoquer des réticences à l'utilisation publique des données, et ainsi restreindre l'utilisation de ces données¹.

Si elle pose les conditions du choix entre les licences ouvertes et ODbL, cette note ne suscitera pas le débat escompté. Lors du Copil suivant, la volonté de privilégier le développement économique par l'ouverture des données détermine le choix de la licence ouverte au détriment de la licence obligeant au partage à l'identique ou d'une licence générique.

Dans un premier temps, nous souhaitons créer le maximum de business à partir de nos données, il faut donc qu'on favorise au maximum leur appropriation. Ensuite, on verra au bout de quelques années, s'il faut revenir sur ce choix et privilégier une licence plus contraignante. Mais, pour le moment, il faut qu'on facilite l'utilisation.²

L'abandon de la licence ODbL au profit de la licence ouverte est acté. Il matérialise l'orientation économique de la stratégie d'ouverture des données.

¹ Source : Note interne « L'ouverture des données au Grand Lyon : Enjeux politiques et économiques du choix de la licence » (10 janvier 2013)

² Source : Journal de terrain, Copil SmartData, Intervention de Nicolas (21 janvier 2013)

C. Une licence « avec authentification » pour s'assurer du respect de l'intérêt général

Sous l'influence de la direction de la voirie, un point important de la stratégie de mise à disposition de données consiste, pour certaines données jugées stratégiques, à authentifier les utilisateurs afin d'être en mesure de contrôler l'usage qu'ils feront des données et de s'assurer que celui-ci soit en cohérence avec la politique de mobilité de la communauté urbaine¹. Pour ce faire, une licence spécifique est rédigée donnant le droit au Grand Lyon de la résilier « pour motif d'intérêt général notamment pour toute exploitation contraire aux politiques publiques mises en œuvre par la Métropole de Lyon ou par le Producteur »².

L'utilisation des données est ainsi soumise au respect de l'intérêt général. Cette référence à un principe civique doit, dans l'esprit des rédacteurs, s'imposer à la liberté entrepreneuriale et commerciale des acteurs privés. Cependant, la rédaction de cette licence va entraîner toute une série de questions aux réponses incertaines. Qu'est-ce que l'intérêt général ? Qui est en mesure de le juger ? Sur quels critères objectifs s'appuyer pour s'assurer de ne pas discriminer les réutilisateurs ?

Le problème que l'on a avec nos licences restrictives, c'est de savoir d'une part, quels sont les critères qui nous permettent de restreindre les réutilisations des données. "Le respect de nos politiques publiques", c'est assez vague, même si nos politiques publiques sont clairement définis quelque part. D'autant plus, que les données des directions ressources ne sont pas concernées par les politiques publiques, puisqu'elles ne mettent pas en place des politiques publiques ! D'autre part, même si l'on arrive à décrire précisément les critères d'utilisation, qui va se charger de les contrôler ? On les contrôle a priori ou a posteriori ? Qui est en mesure de juger ? Quelle est la capacité du Grand Lyon pour contrôler les réutilisations ?³

Les politiques publiques sont définies dans les statuts de la communauté urbaine de Lyon. Néanmoins toutes les données ne peuvent entrer dans ce cadre d'usage restrictif, « les directions devront également clairement démontrer le motif d'intérêt général qui prescrit le placement de la donnée sous cette licence avec authentification. Toutes les données ne pourront donc être placées sous cette licence, il faut qu'il y ait de vraies raisons⁴. » Par exemple, concernant les données mobilité, le référentiel de hiérarchisation des voies et de leur usage autorisé peut permettre de caractériser un manquement au respect de l'intérêt général.

¹ Source : Note interne « Stratégie de mise à disposition des données » (19 mars 2012)

² Source : Licence Engagée de la Métropole de Lyon – Article 17.2

³ Source : Journal de terrain, Discussion avec Clément (4 février 2013)

⁴ Source : Journal de terrain (6 février 2013)

Pour cela, le licencié doit remplir une déclaration de services préalable à l'accès aux données dans laquelle « *il indique la destination des données et les conditions dans lesquelles il projette de les réutiliser ; il décrit notamment les services ou produits dans lesquels les données sont destinées à être intégrées* »¹. L'applicabilité de cette clause est cependant délicate à négocier pour le Grand Lyon. En effet, la communauté urbaine doit appliquer le principe d'égalité des usagers et les conditions de réutilisation doivent être équitables, proportionnées et non discriminatoires². La déclaration de services ne peut ainsi conduire à une sélection a priori des réutilisateurs, le Grand Lyon peut uniquement demander des précisions quant à l'usage qui sera fait des données.

Néanmoins, cette déclaration de service est un moyen pour la communauté urbaine de contrôler a posteriori que l'usage des données est conforme à la déclaration. La licence lui octroie en effet un droit d'audit du système d'information du Licencié. Cependant, malgré ces clauses, l'application de cette capacité de contrôle par le Grand Lyon est délicate. D'une part, parce que chaque direction productrice de données devra mettre en place une instance de contrôle de l'utilisation de ses données. D'autre part, parce que la chaîne de la donnée peut être longue et le signataire de la licence n'est pas toujours l'utilisateur final de la donnée.

Infomobile fournit les données sur l'information mobilité, mais ensuite, ce n'est pas lui qui les intègre dans le calculateur d'itinéraire. Il est donc impossible pour eux de s'assurer que le trafic n'est pas reporté vers des voies secondaires, puisque ce sont d'autres acteurs qui produisent le calcul d'itinéraire. Le constat est le même pour l'ensemble des GPS. Les constructeurs intègrent la hiérarchie des voies, et privilégient les voies 1, mais quand celles-ci sont bouchées, ils reportent automatiquement vers les voies 2 et ainsi de suite³.

Ces articles, présents dans les licences « engagée » et « associée » de la communauté urbaine, contreviennent aux principes initiaux de l'*open data*. L'usage des données n'est en effet pas libre, l'institution se gardant un droit de regard et de contrôle. Cette licence nécessite également d'adapter techniquement l'infrastructure de diffusion de données. Des dispositifs d'authentification des utilisateurs sont indispensables pour limiter l'accès aux données et être en capacité de « couper le flux » en cas de mauvais usage.

¹ Source : Licence Engagée de la Métropole de Lyon – Article 8.2

² Article 38 du décret n°2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour l'application de la loi n°78-753 du 17 juillet 1978

³ Source : Journal de terrain (17 avril 2013)

D'une intention simple de s'assurer du respect des politiques publiques découle toute une série de conséquences mises en lumière lors de la rédaction de cette licence. La volonté de préserver l'intérêt général contribue à traduire l'infrastructure, mais également la communauté urbaine, qui devient une institution garante de l'intérêt général, les données, qui doivent être de nature stratégique, et les utilisateurs, dont on anticipe des usages contraires aux politiques publiques.

D. Une licence avec redevance pour éviter les monopoles

La volonté d'éviter l'appropriation des données ouvertes par un nombre restreint d'entreprises conduit à la rédaction d'une troisième licence. La communauté urbaine souhaite réguler le marché des données ouvertes afin « *d'éviter que des monopoles ne s'installent ou ne viennent fragiliser les modèles économiques fragiles en place sur l'info voyageur¹* ». Pour cela, une licence avec redevance pénalisante pour les acteurs en situation monopolistique est mise en place. Comme pour la licence précédente, ce souhait initial se révèle être un casse-tête juridique autrement plus compliqué et soulève toute une série de conséquences non envisagées. Par une série de tâtonnements, l'équipe en charge de la rédaction compose avec le cadre législatif, les intérêts économiques, les possibilités techniques et les volontés politiques pour parvenir à une licence suffisamment solide.

La législation française autorise les administrations à recevoir une redevance en échange de la mise à disposition de leurs données. Cependant, le produit de cette redevance ne peut dépasser la somme de l'amortissement du coût de collecte et de production de la donnée ainsi que du coût de mise à disposition de cette donnée. En outre, les administrations sont contraintes par un principe de non-discrimination puisqu'elles doivent respecter un principe d'égalité entre les usagers du service public. Néanmoins, ce principe de non-discrimination n'interdit pas les discriminations objectives entre différentes catégories d'usagers. Pour rédiger cette licence « anti-monopole », la communauté urbaine de Lyon doit jouer avec ce cadre légal tout en parvenant à répondre à ses objectifs de développement économique. « *L'objet n'est pas d'éviter que des entreprises puissent construire un business pérenne sur le sujet, au contraire, mais pour éviter que de grands acteurs ne détruisent les rares PME ayant un modèle économique sur le sujet²* ». Toute la difficulté de la rédaction de

¹ Source : Journal de terrain (16 février 2013)

² Source : Journal de terrain (16 février 2013)

cette licence réside dans ces lignes précédentes : comment discriminer les réutilisateurs sur une base objective sans pénaliser les entreprises dont on souhaite favoriser le développement économique?

1) Identifier des critères objectifs de discrimination

Dans le cadre de mes activités de chargé de mission, j'ai été chargé de proposer des premiers critères objectifs de discrimination sur lesquels faire reposer le calcul de la redevance. Comme je le note alors dans mon journal de terrain, l'enjeu est de « *trouver un critère objectif de discrimination permettant de distinguer les « gros » des « petits »* ».

L'objectif du Grand Lyon est de discriminer les réutilisateurs en fonction du bénéfice réalisé par les données afin de « protéger » les « petits » réutilisateurs et d'éviter la formation de monopole. L'entreprise ici visée est Google qui est perçue comme un acteur qui va s'accaparer toutes les données et « étouffer » l'écosystème local. Or, en tant qu'institution publique, le Grand Lyon n'a pas légalement le droit de distinguer les usagers autrement qu'en se basant sur des critères objectifs. Il est donc nécessaire de trouver un critère objectif de discrimination permettant de distinguer les « gros » des « petits »¹.

Après avoir passé en revue la diversité des modèles économiques de la donnée ouverte, j'identifie cinq critères de tarification envisageable.

- Un forfait d'accès à la donnée
- Une tarification proportionnelle à l'usage de la donnée (calculée par le nombre d'appels à l'API)
- Une tarification proportionnelle à la performance (audience), qui peut se calculer par le nombre de visiteurs du site internet, d'utilisateurs du service ou de produits vendus.
- Une tarification proportionnelle à la taille de l'entreprise, déterminée par le chiffre d'affaires global de l'entreprise
- Une tarification proportionnelle à l'exploitation commerciale de la donnée, déterminée par le chiffre d'affaires généré directement ou indirectement (ex. par la publicité) par la réutilisation des données

¹ Source : Journal de terrain (14 mars 2013)

Présentés aux différents membres du comité de pilotage, chacun de ces critères soulève des réserves relevant de problématiques diverses.

- La première proposition soulève des préoccupations économiques et organisationnelles. Le forfait d'accès instaure une barrière à l'entrée pour les réutilisateurs : *« on risque d'empêcher les petits utilisateurs qui souhaitent se lancer, mais qui ne sont pas sûr de gagner de l'argent »*¹. En outre, il présente des coûts de gestion importants pour le Grand Lyon pour des sommes minimales. *« Je préfère que l'on ne facture rien plutôt que de facturer des petites sommes avec des coûts de gestion importants »*².
- La deuxième proposition se heurte à des contraintes techniques. *« Il est techniquement impossible de baser la facturation sur l'usage effectif de la donnée basé sur le nombre d'appels à l'API. D'une part, les entreprises ne font qu'un seul appel à l'API et dupliquent le flux dans leur système d'information. D'autre part, on pourrait leur demander de nous envoyer des jetons pour chacune des utilisations qui sont effectuées ensuite, mais notre architecture n'est pas adaptée pour cela, et cela s'avère complexe à mettre en œuvre »*³.
- La troisième proposition ne couvre qu'une partie des valorisations commerciales des données ouvertes. Les usages commerciaux de la donnée ne donnant pas lieu à la vente de produits ou de services, comme par exemple les usages internes des données, ne sont pas pris en compte. En outre, la tarification proportionnelle à la performance (audience) ne permet pas de *« distinguer les services ou produits destinés à un marché de niche ou de masse. Elle fonctionne très bien pour des services fournis sur internet ou via des applications mobiles destinés à un grand public (cf. Google), mais plus difficilement dans d'autres cas »*⁴.
- Les quatrième et cinquième propositions risquent d'être difficiles à calculer et de susciter des réticences de la part des utilisateurs. Un échange avec une entreprise nous révèle que *« le critère du chiffre d'affaires n'était pas viable. En effet, la ville de Paris avait mis en place une tarification en fonction du CA généré. Mais, c'était très compliqué pour eux de calculer cela, avec tous les modèles économiques existants et les cas particuliers »*. En effet, la chaîne d'usage des données peut s'avérer beaucoup plus longue et complexe qu'envisagée initialement comme nous le révèle un échange avec une entreprise : *« Par exemple, à partir des données Grand Lyon et celles d'autres sources, on produit un flux qui est revendu à des entreprises qui l'intègrent dans leurs systèmes de navigation. [...] Il y a aussi des équipements que l'on ne vend pas directement. On a alors une chaîne comme le Grand Lyon fournit les données à [une entreprise spécialisée en données mobilité], qui les traite pour Michelin Partner,*

¹ Source : Journal de terrain (18 mars 2013)

² Source : Journal de terrain (28 mars 2013) – Echange avec Harald.

³ Source : Journal de terrain (18 mars 2013)

⁴ Source : Journal de terrain (28 mars 2013) – Echange avec Harald.

qui vend le flux à un équipementier, Bouygues par exemple, qui lui même le vend à un constructeur (Peugeot), qui le vend au consommateur final¹. »

- En outre, la cinquième proposition ne répond pas à l'objectif souhaité par le Grand Lyon puisqu'elle ne couvre pas les produits d'appel, fournis gratuitement et financés par des budgets provenant d'autres marchés. *« Il faut gérer le cas où un acteur propose un service gratuit, avec de facto un financement assuré par d'autres services, qui le mettrait en situation de monopole à terme, service gratuit qui pourrait ensuite être financé par la pub ... Le critère CA ne gère pas ce cas, Google map pour en citer un². »*

A l'issue de plusieurs aller-retour et échanges avec des entreprises utilisatrices de données mobilité, deux barèmes de tarification sont finalement adoptés : le nombre de visites mensuelles sur le site internet du service et l'activité générée par la réutilisation des données calculée en fonction du nombre de produits vendus réutilisant les données de la Communauté urbaine de Lyon. Ces échanges manifestent la complexité de la mise en œuvre d'une idée initiale claire : un état du marché doit être anticipé et la viabilité technique, économique, organisationnelle doit être questionnée. L'étape suivante de détermination du montant de la redevance fait intervenir tout autant d'intérêts hétérogènes.

2) Déterminer le montant de la redevance

La détermination du montant de la redevance est un exercice délicat. La législation impose que la redevance soit d'un « montant raisonnable ». Personne n'est cependant en mesure de préciser ce qu'est un montant raisonnable. Quoi qu'il en soit, il ne peut dépasser l'amortissement du coût de la production et celui de la mise à disposition de la donnée. Cette incertitude nous oblige à tâtonner pour déterminer un montant qui soit suffisamment dissuasif pour éviter la réutilisation par certains « gros » acteurs, sans toutefois pénaliser les entreprises dont on souhaite encourager le développement économique, ni « *les grands groupes, qui sont les principaux moteurs de l'innovation sur le territoire et qui sont nos partenaires³* ».

Pour s'assurer de leur propension à payer, une série de rencontres avec des entreprises partenaires assignées porte-parole des utilisateurs de données publiques est organisée. Lors de

¹ Source : Journal de terrain (10 juin 2013)

² Source : Journal de terrain (28 mars 2013) – Echange avec Harald.

³ Source : Journal de terrain (28 mars 2013)

ces échanges, une entreprise s'inquiète du montant de la redevance initialement proposée par le Grand Lyon qui la contraindrait à ne plus utiliser ces données.

Comme je vous le disais tout à l'heure, la situation économique et le marché ne sont pas au beau fixe. Donc, quand PR m'a dit que vous souhaitiez passer d'une redevance de 500 euros aujourd'hui à 25 000 demain, ça m'a fait réagir, et j'ai voulu vous voir. Nous, aujourd'hui, 25 000 euros, on ne peut pas les mettre¹.

Le montant de la redevance sera adapté afin que les entreprises de taille similaire à celle-ci ne soient plus concernées par les redevances. Une tarification proportionnelle est mise en œuvre, mais l'incertitude persiste sur le montant maximal que l'on peut afficher.

Antoine : Je dois mettre quels montants pour chaque seuil ?

Harald : Je sais pas... Si on met un montant maximal à 300 000€, on va avoir des articles dans la presse qui vont dire : « Le GL vend ses données 300 000€ ». Mais, de toute façon, il s'agit d'un affichage pour éviter que Google vienne utiliser les données. Personne ne paiera cette somme maximale².

Sans véritable étude complémentaire, les seuils de redevance seront déterminés de manière approximative. Les seuils et le montant de la redevance indiqués dans l'annexe de la licence sont les suivants³ :

Nombre de visites mensuelles	Volume d'activité	Redevance annuelle (non délibérée)
< 300 000	< 40%	0 €
De 300 000 à 600 000	De 40 à 50%	50 000 €
De 600 001 à 1 000 000	De 50 à 75%	150 000 €
> 1 000 000	> 75%	450 000 €

*

**

Ces quelques points de débat soulignent l'ensemble des actants qui interviennent dans la détermination d'un critère de redevance : l'infrastructure technique de diffusion, les régies publicitaires d'Internet, l'entreprise Google et son modèle économique, les données, les modèles économiques des données ouvertes, la population du Grand Lyon, les PME locales,

¹ Source : Journal de terrain (10 juin 2013)

² Source : Journal de terrain (28 mars 2013) – Echange avec Harald.

³ Source : Annexe 3 de la Licence associée du Grand Lyon. http://data.grandlyon.com/files/2015/04/Annexe-3_redevances_LicenceAssociee_Ch.pdf (Consulté le 21 avril 2016)

des variables micro-économiques (chiffre d'affaires, part de marché), l'organisation interne de la communauté urbaine et le coût de mise en place d'une régie de données, les relations du Grand Lyon avec certains acteurs économiques, le risque politique pour les élus, la presse, etc. Le choix de ces critères de redevance repose également sur des hypothèses, par nature incertaines, qui soulèvent de nombreuses questions : l'entreprise Google sera-t-elle réfractaire au montant de la redevance ? Est-ce que l'on couvre tous les modèles économiques basés sur les données ouvertes ? Peut-on déterminer les parts de marché de ces services ? Est-ce que le critère de proportionnalité reposant sur les parts de marché est équitable ? Peut-on isoler le chiffre d'affaires de l'activité réalisée par l'usage des données ouvertes ? Comment distinguer les marchés de masse des marchés de niche ?

Au-delà des critères finalement adoptés, la présentation de ce processus de création d'une licence avec redevance met en lumière l'incertitude dans laquelle sont placés les acteurs de l'administration. A partir d'un cadre légal et d'un objectif initial, ils procèdent par tâtonnements, par allers-retours, par échanges avec différents acteurs pour réussir à proposer un modèle de licence avec redevance qui tienne. La solidité de cette licence repose sur les différentes traductions successives qui ont permis d'y associer des actants divers : données, infrastructures, modèles économiques, services juridiques, producteurs de données, utilisateurs de données, etc. ou au contraire d'en maintenir d'autres à l'écart.

L'authentification et le contrôle deviennent une démarche partenariale visant un engagement entre le Grand Lyon et les entreprises utilisatrices. Le paiement d'un coût fixe d'accès aux données est abandonné, et la licence devient gratuite par défaut, payante en cas de situation de redevance. Ces modifications peuvent sembler n'être que des éléments de langage visant à promouvoir le Grand Lyon. Pourtant, elles transforment l'identité des acteurs en présence. La communauté urbaine de Lyon n'est plus dans une posture de contrôle, mais dans une logique de régulation. Les entreprises ne sont plus des acteurs commerciaux autonomes, elles sont dorénavant des partenaires de l'institution publique, qui, dans ce cadre, participent à l'action publique. Les données ne sont plus une ressource patrimoniale que le Grand Lyon peut valoriser économiquement : elles n'ont plus de valeur en soi, mais n'en acquerront que dans l'usage qui en sera fait (on passe d'une logique de stock à une logique de flux). Le travail sur les critères de redevance clarifie la valeur attribuée aux données par le Grand Lyon. *« L'objectif est de ne pas pénaliser les PME et au contraire d'éviter que des grands tuent les*

petits. Il ne s'agit pas de faire de l'argent avec les données. » Les redevances ne sont plus une manne financière pour la collectivité, mais un instrument de régulation.

*

* *

En résumé, ce travail d'instrumentation se traduit par trois licences qui formalisent les grands principes de la politique de diffusion des données de la communauté urbaine de Lyon. La licence ouverte doit favoriser les réutilisations commerciales des données ; la licence avec authentification vise à assurer le Grand Lyon du respect de ses politiques publiques ; et la licence avec redevance à empêcher la constitution de monopoles. Adoptées à l'unanimité par le conseil communautaire sans aucun débat¹, ces licences entérinent le travail de problématisation de la politique d'ouverture des données, pensée et définie par l'administration.

IV. Conclusion – Pluralisme des politiques d'open data, pluralisme du rôle du grand lyon

La généalogie de la politique d'ouverture des données de la communauté urbaine de Lyon met en avant la diversité des cadrages qui s'affrontent au sein de multiples arènes. La problématisation est une activité de composition, au sens qu'en donne Bruno Latour : un travail de rassemblement, élément après élément, d'intérêts, de valeurs et d'entités incommensurables dans un *modus vivendi* commun, toujours fragile et révisable². La politique d'ouverture de données qui résulte de ce travail de composition est le résultat de traductions successives qui agrègent et alignent les acteurs pluriels autour d'une version partagée du problème public. Plutôt que de les considérer de manière uniforme, les politiques d'ouverture de données doivent ainsi être envisagées au pluriel en prenant en compte les traductions locales assignées à chacune d'entre elles.

¹ L'absence de débat et le vote à l'unanimité ne sont pas propres à l'*open data*, mais sont une constante des institutions intercommunales, comme le démontrent Desage et Guéranger. Fabien Desage et David Guéranger, *La politique confisquée : Sociologie des réformes et des institutions intercommunales*, Bellecombe-en-Bauges, Editions du Croquant, 2011, 247 p.

² Il s'agit d'ailleurs de la définition de ce qu'est la *politics* pour Latour, « *the progressive composition of a common world* », qui pointe au travers de cette définition le nécessaire effort de construction d'un espace partagé. B. Latour, *Reassembling the Social. An introduction to Actor-Network Theory*, *op. cit.*, p. 254.

La problématisation de la politique d'ouverture de données du Grand Lyon est le résultat d'une succession d'opérations de traduction. D'un univers des possibles qui est très ouvert, on le ferme petit à petit pour stabiliser la problématisation de la politique d'ouverture des données. On polit le projet pour lui enlever toutes les aspérités et en faire quelque chose qui tient, c'est-à-dire qui n'est pas remis en question et dont la chaîne des associations est suffisamment solide. On a retiré des associations jugées insuffisamment solides (transparence, participation) pour se focaliser sur certains aspects (développement économique). Ces différentes épreuves ne redéfinissent pas seulement la politique d'ouverture des données, mais également l'ensemble des actants associés à cette politique : les données, les utilisateurs et même la communauté urbaine.

Au cours de ce processus de problématisation, les données sont toujours les mêmes, mais elles sont catégorisées de multiples façons selon les récits dans lesquels elles sont insérées. La donnée est ainsi successivement une information géographique, un outil au service des politiques publiques, une information pour l'utilisateur, une prise pour la transparence et la participation citoyenne, un levier de développement économique, etc.

Les utilisateurs des données ouvertes prennent également de multiples formes. Ils sont tour à tour les communes de la communauté urbaine, les acteurs partenaires de l'action publique territoriale, les associations citoyennes, les usagers et les entreprises.

Enfin, la communauté urbaine elle-même voit son identité d'action être transformée selon les différentes problématizations, sans que ces différentes missions qui lui sont assignées ne soient véritablement mises en débat. Alors qu'on lui assigne un rôle de fournisseur de service pour les communes de son territoire, elle devient peu à peu le responsable de la politique de mobilité de l'agglomération, un acteur public garant de l'intérêt général, l'animateur du développement économique, le promoteur d'une métropolisation par les données avec la plateforme d'agglomération, etc. Par l'ensemble de ces rôles, la communauté urbaine assume un rôle de coordination de l'ensemble des acteurs urbains au travers des données mises en circulation.

Chapitre 3

Ouvrir : Les données à l'épreuve de leur diffusibilité

En février 2015, alors que j'ai débuté ma thèse sur l'ouverture des données depuis plus de trois ans, lors d'une réunion, Karine Dognin-Sauze, vice-présidente aux nouvelles technologies de la Métropole de Lyon me pose une question en apparence simple, mais qui se révèle assez embarrassante : « *qu'est-ce qu'une donnée ?* ».

Karine Dognin-Sauze m'a posé cette question lors de notre réunion : « *Pour vous, c'est quoi une donnée ? Comment on distingue ce qu'est une donnée ?* » J'ai bafouillé une réponse à cette question, en reprenant dans un premier temps la distinction entre données, information et connaissance, avec l'interprétation de la donnée qui en fait une information. Puis, pour répondre de manière beaucoup plus pragmatique, j'ai ajouté qu'une donnée était ce qui pouvait être traité informatiquement, notamment tout ce qu'on pouvait mettre sous la forme de fichiers excel par exemple. La réponse était peu satisfaisante, et elle l'a bien senti¹.

Mon embarras pour répondre à cette question est révélateur de la difficulté à définir ce qu'est une donnée. Alors que les sciences de l'information², l'informatique³ ou le droit donnent des définitions variées de ce que sont les données, ces catégorisations ne sont que très rarement mobilisées en situation. A l'exception des juristes de l'administration, pas une seule fois lors de la mise en place de la politique les acteurs ne feront référence à la distinction entre données et informations, ou à la catégorisation juridique d' « information publique »⁴.

Si je reviens sur cette anecdote, c'est que le travail de catégorisation est l'enjeu central du processus d'ouverture de données qui est l'objet de ce chapitre. Ouvrir une donnée nécessite en premier lieu d'identifier ce qu'est une donnée, puis de déterminer sa diffusibilité. Pour comprendre comment une donnée devient une donnée ouverte, il faut observer, en situation, ce que les acteurs entendent par le terme de « données », de « données publiques », de « données ouvertes », de « données candidates », de « données personnelles », de « données sensibles », etc. Ce travail de catégorisation est au cœur de la politique d'ouverture de données : plus que d'éventuelles résistances d'acteurs accrochés à leur pouvoir, il détermine pourquoi certaines données sont, ou non, mises à disposition d'un nouveau public.

¹ Source : Journal de terrain – Réunion avec Karine Dognin-Sauze (18 février 2015)

² En science de l'information, la donnée est distinguée de l'information et de la connaissance, ces trois entités étant organisées hiérarchiquement : la donnée préexiste, puis l'information et enfin la connaissance. La donnée est définie comme la description élémentaire d'une réalité, l'information comme une donnée qui acquiert du sens, et la connaissance comme une interprétation des informations.

³ Si l'on s'en tient à une conception informatique, la donnée est un élément binaire, représenté par une séquence de *bits*, qui peut circuler, être stockée et être analysée.

⁴ La législation française ne fait pas référence à la notion de données publiques, mais à celles de « documents administratifs » et « d'informations publiques ».

Les « barrières » qui restreindraient l'ouverture des données occupent une place centrale dans les écrits, militants ou académiques, sur l'*open data*¹. Janssen et al. pointent le fait que l'ouverture des données implique une transformation institutionnelle d'un système fermé à un système ouvert². Gray et Davies appellent à passer d'une vision de la « libération » des données à une politique de recomposition de l'infrastructure informationnelle³. Les barrières à lever pour permettre ce changement sont nombreuses : héritage institutionnel, emprise politique, aversion au risque, complexité de la démarche, contraintes techniques, incertitude juridique, culture « fermée » des administrations, etc. A partir de l'étude de politiques locales d'*open data*, Peter Conradie et Sunil Choenni pointent trois principaux facteurs limitant la mise à disposition des données : le stockage décentralisé des données, les sources externes de données, le non-usage de la donnée dans le cœur du service public⁴.

En soulignant les nécessaires recompositions institutionnelles, cette littérature pointe le fait que les données ne sont pas autonomes d'un environnement social. Néanmoins, en se focalisant sur les « barrières » de l'*open data*⁵, ces auteurs s'abstiennent de s'intéresser aux données elles-mêmes. Dans leur perspective, les données préexistent à l'ouverture et elles ne jouent aucun rôle dans ce processus dans la mesure où il suffit de convaincre leurs propriétaires de lever des « barrières » pour les diffuser. Or, l'ouverture des données redéfinit tout autant les données que l'environnement dans lequel elles sont insérées. Pour comprendre comment une « donnée fermée » devient une « donnée ouverte », il est nécessaire d'adopter une perspective relationnelle qui étudie symétriquement les données et l'environnement dans lequel elles sont insérées.

¹ Chris Martin, « Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective », *Policy & Internet*, 2014, vol. 6, n° 3.

² Marijn Janssen, Yannis Charalabidis et Anneke Zuiderwijk, « Benefits, Adoption Barriers and Myths of Open Data and Open Government », *Information Systems Management (ISM)*, vol. 29, n° 4, p. 258-268.

³ *Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures?*, 2015.

⁴ Peter Conradie et Sunil Choenni, « On the barriers for local government releasing open data », *Government Information Quarterly*, 2014, vol. 31, p. S10-S17.

⁵ En se focalisant sur les « barrières », ces auteurs s'inscrivent dans le registre des militants de l'*open data* qui prônent la « libération » des données par « l'ouverture » des institutions. Samuel Goëta a montré comment ce discours se manifestait dans l'iconographie associée à l'*open data*.

Samuel Goëta, *Open data : qu'ouvre-t-on avec les données publiques ?*, Université de Paris IV - Sorbonne, Paris, 2011, 101 p.

Un des impensés de ces discours est qu'en se focalisant sur les institutions, ils ne prennent pas en compte les données elles-mêmes qui sont constitutives des liens qui les attachent (voir Chapitre 1).

Dans une perspective sociotechnique, je propose de m'intéresser au processus de définition de la donnée en situation au travers des épreuves successives au cours desquelles elle est questionnée. La donnée subit en effet une série d'épreuves qui détermine son avenir en définissant sa « diffusibilité¹ ». Forcée par la sociologie pragmatique des sciences et des techniques, l'épreuve est une situation d'incertitude au cours de laquelle l'identité des êtres est questionnée. Elle est particulièrement adaptée pour analyser les catégorisations successives dont sont l'objet les données au cours du processus préalable à leur diffusion avant d'acquiescer la définition stabilisée de « donnée ouverte ». L'épreuve ne se contente pas de questionner les données, mais elle réinterroge l'ensemble des actants de cette situation : producteurs, infrastructure technique, acteur public, utilisateur, etc.. Ce moment de redéfinition généralisée permet d'aligner les intérêts de ces différents actants afin d'envisager l'insertion de la donnée au sein d'un autre réseau sociotechnique. Si l'épreuve s'achève par une stabilisation des qualités et attributions des êtres, ce résultat est toujours incertain et provisoire. Les conditions de félicité de la mise à disposition d'une donnée ne sont fixes et stables que jusqu'à la prochaine épreuve. Une donnée qui n'est pas diffusible peut le devenir (et vice-versa) si sa catégorisation est remise en question.

En catégorisant différemment la donnée, l'épreuve participe au détachement de la donnée de l'infrastructure informationnelle dans laquelle elle est insérée. En effet, la donnée n'est jamais brute², mais elle est toujours étroitement associée à un ensemble de personnes, de pratiques, de technologies, d'institutions qui les produisent, les maintiennent et les utilisent³ (cf. chapitre 1). Pour permettre sa diffusion et son utilisation dans un environnement autre, il est nécessaire de délier l'ensemble de ces attachements afin de rendre la donnée autonome de ce cadre initial. Cela oblige à reprendre une à une les composantes de l'attachement, ce qui exige de prendre en compte la donnée dans toutes ses dimensions : juridiques, techniques, économiques, politiques, etc.

¹ La diffusibilité est un terme scientifique caractérisant l'aptitude d'une substance fluide à se diffuser (gaz, lumière). Ce terme est préféré à celui, plus courant, de diffusabilité, utilisé pour désigner quelque chose que l'on peut diffuser. En effet, il est plus adapté à l'argument de ce chapitre puisqu'il souligne le fait que la donnée est transformée au cours du processus d'ouverture pour acquiescer des propriétés de diffusibilité.

² Geoffrey C. Bowker, « Biodiversity Datadiversity. », *Social Studies of Science*, 2000, vol. 30, n° 5, p. 643-683 ; Lisa Gitelman (ed.), « *Raw Data* » *Is an Oxymoron*, Cambridge Mass., MIT Press, 2013, 192 p.

³ Paul Edwards et al., « Introduction: An Agenda for Infrastructure Studies », *Journal of the Association for Information Systems*, 2009, vol. 10, n° 5, p. 364-374 ; Geoffrey C. Bowker et al., « Toward information infrastructure studies: ways of knowing in a networked environment » dans Jeremy Hunsinger, Lisbeth Klastrup et Matthew Allen (eds.), *International Handbook of Internet Research*, New York, Springer, 2010, p. 97-118.

Le travail de détachement est toutefois indissociable d'un travail d'attachement. Dissocier, c'est créer de nouveaux liens autant que d'en défaire d'autres¹. Pour susciter l'intérêt à l'*open data*, la donnée doit être associée à de nouveaux enjeux (développement économique, émergence de nouveaux services, simplification du travail des agents, image innovante de la collectivité, etc.) et à de nouveaux utilisateurs (entreprises, développeurs, citoyens, agents de collectivités territoriales, etc.). Tout autant que la donnée, le public des données ouvertes est défini au travers de ces épreuves. Des usages potentiels des données sont préfigurés tout au long du processus, un travail de cadrage², bien connu des sociologues des techniques, que Madeleine Akrich a défini comme des scripts³. Etudier le processus de diffusion des données c'est analyser symétriquement les médiations qui détachent et les médiations qui attachent.

Ainsi, le travail préliminaire à la mise à disposition des données éprouve à la fois le détachement des données de leur infrastructure informationnelle et leur attachement à de nouveaux utilisateurs. Multiples, ces épreuves de *diffusibilité* sont regroupées en trois catégories : l'identification, la publicisation et l'extraction.

¹ Ce point est présent dans le texte fondateur de la sociologie de la traduction de Michel Callon. Michel Callon, « Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques dans la Baie de Saint-Brieuc », *L'Année sociologique*, 1986, n° 36 ; Dans la sociologie de l'innovation, l'activité de détachement est néanmoins souvent négligée au profit de l'analyse de l'activité d'attachement. A partir du cas des innovations « par retrait », Goulet et Vinck nous invitent à analyser symétriquement ces deux processus. Frédéric Goulet et Dominique Vinck, « L'innovation par retrait. Contribution à une sociologie du détachement. », *Revue française de sociologie*, 2012, vol. 53, n° 2, p. 195-224 ; Dans son travail sur les addictions, Gomart souligne que le détachement des individus aux drogues passe par leur attachement à de nouvelles substances substitués des précédentes. Emilie Gomart, *Surprised by Methadone*, Ecole des Mines, Paris, 1999 ; Goulet et Le Velly mettent en avant un processus similaire dans le cadre des échanges marchands où l'attachement à un nouveau produit nécessite le détachement du précédent. Frédéric Goulet et Ronan Le Velly, « Comment vendre un produit incertain ? Activités de détachement et d'attachement d'une firme d'agrofourmiture », *Sociologie du Travail*, 2013, vol. 55, n° 3, p. 369-386 ; Enfin, Laurent Fouillé démontre que le détachement automobile n'est réalisable que par attachement à des modes de transport alternatifs. Laurent Fouillé, *L'attachement automobile mis à l'épreuve. Etude des dispositifs de détachement et de recomposition des mobilités.*, Université Rennes 2, Rennes, 2010, 476 p.

² E. Goffman, *Les cadres de l'expérience*, op. cit.

³ Madeleine Akrich, « Comment décrire les objets techniques ? », *Techniques et culture*, 1987, n° 9, p. 49-64.

I. Identifier

La donnée ne préexiste pas à son ouverture : elle n'est pas déjà-là, prête à être mise à disposition. Comme le soulignent Jérôme Denis et Samuel Goëta, les données publiques « *ne sont pas disponibles en l'état, prêtes à être libérées. Leur existence même est loin d'être une évidence*¹ ». Pour ouvrir une donnée, il faut en premier lieu savoir ce qu'est une donnée, quelles données existent et qui les détient. L'identification est la première épreuve de diffusibilité des données. Au travers d'une exploration au sein des services de la Métropole, elle consiste à désigner une entité comme une donnée.

A. Explorer l'espace des données de l'agglomération lyonnaise

Identifier les données candidates à l'ouverture n'est pas une tâche aisée. Une institution telle que la communauté urbaine ne connaît pas son patrimoine de données de manière exhaustive. Il n'existe pas en son sein de catalogue² recensant l'ensemble des données traitées par les différents services et directions³. Dès lors, comme l'affirme un urbaniste du système d'information : « *aujourd'hui, on ne sait pas ce que l'on possède*⁴ ». En effet, contrairement aux applications, les données n'ont pendant longtemps pas été perçues comme un actif stratégique du système d'information.

La donnée de gestion n'a jamais été une valeur en soi. Elles étaient cachées dans le système, encapsulées dans les applications. Ce sont les applications qui avaient une valeur, pas les données. Les données n'étaient que ce qui permettait de faire fonctionner l'application⁵.

Les services informatiques de la communauté urbaine connaissent l'ensemble des applications informatiques, mais pas les données que celles-ci produisent ou traitent quotidiennement. Les agents utilisent des données pour réaliser leurs missions quotidiennes,

¹ Jérôme Denis et Samuel Goëta, « La fabrique des données brutes. Le travail en coulisses de l'open data. » dans Clément Mabi, Jean-Christophe Plantin et Laurence Monnoyer-Smith (eds.), *Penser l'écosystème des données. Les enjeux scientifiques et politiques des données numériques.*, Paris, Editions de la Maison des Sciences de l'Homme, 2015, .

² La loi CADA impose pourtant aux administrations un catalogue de leurs informations publiques. « *Les administrations qui produisent ou détiennent des informations publiques tiennent à la disposition des usagers un répertoire des principaux documents dans lesquels ces informations figurent.* » (Article 17 de la loi n°78-753)

³ Ce projet de catalogage exhaustif des données de l'institution est régulièrement annoncé comme indispensable par les acteurs au sein du Grand Lyon, mais également des villes américaines étudiées (New York, Philadelphie, Chicago).

⁴ Source : Journal de terrain (25 mars 2014)

⁵ Source : Entretien avec Nicolas, directeur de la DSIT, communauté urbaine de Lyon (16 janvier 2013)

sans qu'il n'y ait de centralisation de la connaissance de la production, l'acquisition ou l'usage de ces données. L'espace informationnel constitué par les données du Grand Lyon est ainsi fragmenté en différents silos, dont les relations sont faibles ou inexistantes. A mon arrivée à la communauté urbaine de Lyon, je m'étonne de cette fragmentation métiers des données : « *en interne, on ne parle pas des données du Grand Lyon, mais des données de telle ou telle direction* »¹.

Les agents des services informatiques ne sont pas dupes de cette segmentation en silos qui leur empêchent d'avoir une connaissance centralisée des données du système d'information. Lors d'un projet – avorté - de catalogage des données, ils font le constat d'une quadruple segmentation des données qui reflète la multiplicité des mondes sociaux : par silo « métier » (environnement, développement économique, aménagement du territoire), par silo « technico-culturel » (géographie, document, patrimoine), par silo « applicatif » (une application se décline en front office, middle office, back office et business intelligence) et par silo « technique » (Microsoft, IBM, Oracle, propriétaire, open source, etc.)². La multiplication des silos entraîne une méconnaissance du patrimoine informationnel : à la fois son existence (définition, source, disponibilité), sa qualité (exhaustivité, conformité, complétude, duplication) et sa gouvernance (propriété, pérennité, confidentialité)³. Cette situation est problématique alors que, sous l'impulsion de l'*open data*, les données sont maintenant perçues comme un patrimoine informationnel qu'il est nécessaire de connaître pour pouvoir le valoriser.

Aujourd'hui, et le mouvement de l'*open data* a permis de pointer cela, on se rend compte que la donnée a une valeur en elle-même. Elle représente un patrimoine, et à ce titre, il faut le gérer comme tel, en prenant soin de l'entretenir et de le défendre. Donc, le système d'information passe des applications aux données elles-mêmes⁴ !

A l'origine du projet *open data*, les acteurs en charge de l'ouverture des données ignorent l'étendue de ce patrimoine. « *Quand on a commencé à bosser sur le projet, on n'avait aucune idée de quelles données on [la DSIT de la communauté urbaine] gérait* »⁵. Les chefs de projet se mettent dans une posture exploratoire leur permettant de clarifier et

¹ Source : Journal de terrain (17 décembre 2012)

² Source : Document de travail – Projet MADO (Novembre 2010)

³ Source : Document de travail – Projet MADO (Novembre 2010)

⁴ Source : Entretien avec Nicolas, directeur de la DSIT, communauté urbaine de Lyon (16 janvier 2013)

⁵ Source : Journal de terrain, François (24 mars 2014)

d'ordonner un espace dans lequel ils tentent de s'orienter : quelles données sont produites par le Grand Lyon ? Qui gère quoi ? Qui est responsable de quoi ?

A la manière d'explorateurs, ils s'appuient sur différents outils, qu'ils perfectionnent peu à peu, pour tenter de s'orienter au sein de l'espace informationnel de l'agglomération : des cartes, des acteurs « référents » au sein des institutions, des réseaux de producteurs de données, la connaissance d'utilisateurs externes, etc. Ces outils leur offrent différentes pistes qu'ils vont suivre de manière décousue afin d'identifier les données candidates à l'ouverture. Comme lors de toute démarche exploratoire, l'incertitude règne, et les découvertes sont parfois le fruit du hasard de la sérendipité¹. Ce travail exploratoire ne se déroule pas uniquement au début du projet, mais il est permanent et toujours actualisé afin de continuer à enrichir le portail de mise à disposition de données. Accompagnons les deux chefs de projet dans leur voyage exploratoire de l'espace informationnel métropolitain.

1) La carte et le périmètre des « données candidates à l'ouverture »

Le premier outil que les chefs de projet utilisent est une cartographie théorique du patrimoine informationnel de la communauté urbaine. Cette cartographie représente sommairement le patrimoine informationnel : des « réalités » diverses sont « capturées » sous forme de « contenus », structurés ou non structurés, qui sont ensuite inscrits dans des « documents ». A partir de cette représentation du patrimoine informationnel communautaire, est délimité un périmètre de « données candidates à l'ouverture ».

¹ Nicolas Auray, « Les technologies de l'information et le régime exploratoire » dans Pek Van Andel et Bourcier Danièle (eds.), *La sérendipité dans les arts, les sciences et la décision*, Éditions Hermann., Paris, 2010, .

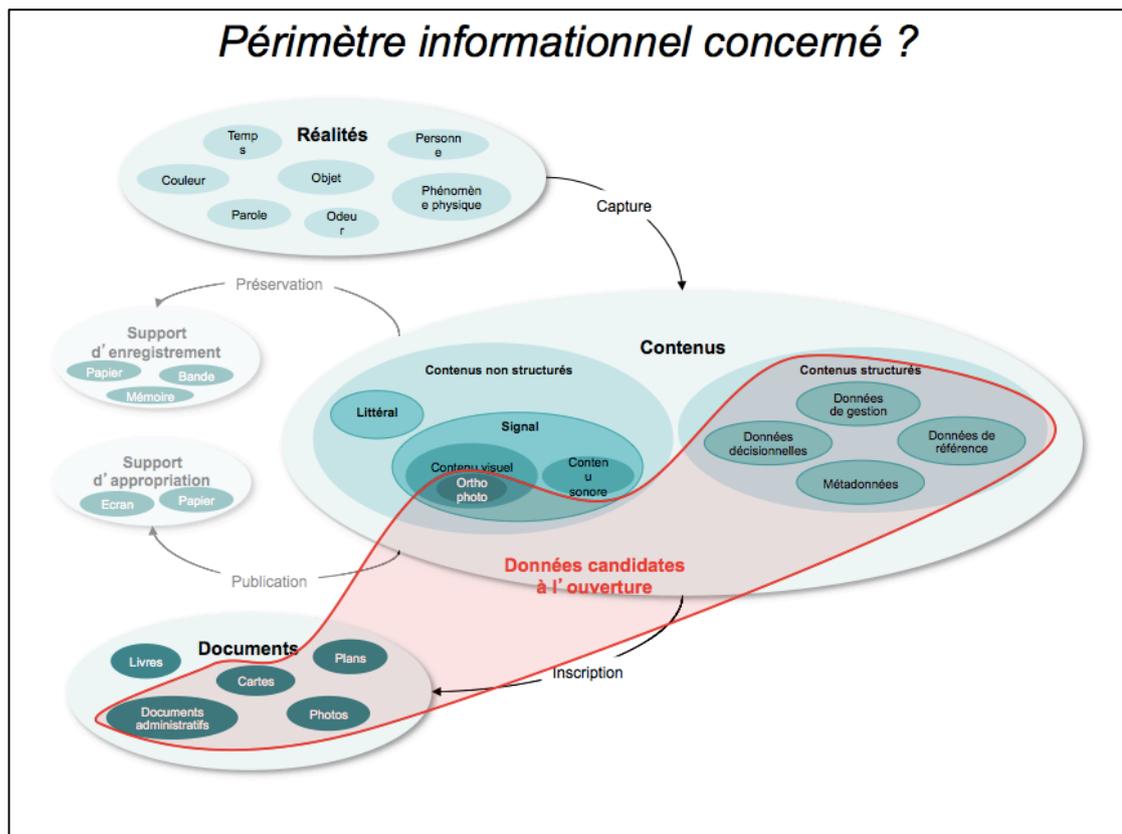


Illustration 17 : Cartographie théorique du périmètre informationnel concerné par l'open data¹

Ce périmètre initial définit les « données candidates à l'ouverture » de manière très large. On y retrouve pêle-mêle des données de gestion et de référence, des métadonnées, des orthophotographies, des documents administratifs, des cartes, des plans, des photos, etc. Cette première cartographie a le mérite d'identifier a priori des données susceptibles d'être mises à disposition. Les renseignements sur l'identité de ces données sont cependant très vagues : on ne sait pas ce qu'elles représentent, qui les détient ou les utilise.

2) Rencontrer les administrateurs de données géographiques

Pour approfondir cette exploration, les chefs de projet rencontrent les administrateurs de données géographiques du Grand Lyon. Ces individus sont les responsables des systèmes d'information géographique de chacune des directions du Grand Lyon. Par chance pour nos chefs de projet, ces administrateurs sont constitués en réseau et ils disposent de listes précises

¹ Source : Document interne « Management des données. Naissance d'un nouveau programme ? » (novembre 2010)

et exhaustives des données qu'ils gèrent. En une réunion, ils peuvent toucher toutes les données géographiques de toutes les directions de la communauté urbaine.

En janvier 2012, François leur présente le projet d'ouverture des données publiques et leur demande de réaliser un état des lieux des données dont ils disposent, d'évaluer pour chacune d'entre elles la volonté de diffusion et d'indiquer un degré de priorisation de leur ouverture (immédiate, moyen terme, long terme)¹.

Liste des couches "données de référence" pour diffusion sur portail cartographique et en "open data"																
Thème	Couche	Portail de visualisation / téléchargement					Site Open Data					Géonet				
		Visualisable en service WFS ?	Téléchargeable sur portail ?	Priorité (du + au - "urgent")	Données disponibles ?	Métadonnées disponibles ?	Lot Open Data "sur étiquette" : préciser étiquette et format	L2	CC 46	Priorité (du + au "urgent")	lots de données disponibles	Métadonnées disponibles	Commentaires	Déjà présent	A ajouter	
Collecte	circuit de collecte	non	non	3	oui	non									Non	Non
	silos verre	oui	oui	1	oui	oui									Non	Oui
	silos emballés	oui	oui	1	oui	oui									Non	Oui
	producteur	non	non	0	oui	non									Non	Non
	points de collecte	non	non	0	oui	non									Non	Non
	centres de tri et usines d'incinération	oui	oui	1	oui	oui									Non	Oui
	déchèteries	oui	oui	1	oui	oui									Non	Oui
	tranchées forains	oui	oui	1	oui	oui									Non	Oui
	poubelles de rue	oui	oui	1	oui	oui									Non	Oui
	bois de la DP	non	non	2	oui	oui									Non	Oui
	toilettes publiques (GL)	oui	oui	1	oui	oui									Non	Oui
	circuits de nettoyage	non	non	0	oui	non									Non	Non
	parcours de nettoyage	non	non	0	oui	non									Non	Non
	parcours de déchèterie	non	non	0	oui	non									Non	Oui
	parcours de fauchage	non	non	0	oui	non									Non	Oui
	parcours de cantonnement	non	non	0	oui	non									Non	Non
espaces canins	non	non	1	oui	oui									Non	Oui	
espaces conventionnés (CQP à ferme)	non	non	3	oui	oui									Non	Oui	
VH	circuits de VH	oui	oui	1	oui	oui									Non	Oui
	coffres à sel	non	non	0	oui	non									Non	Oui
	dépôts de sel	non	non	0	oui	non									Non	Oui
Distinction à prévoir entre GD PUBLIC et COMMUNES																
Collecte		Communes		Grand public												
	silos verre	oui	oui													
	silos emballés	oui	oui													
	centres de tri et usines d'incinération (inspire annexe III)	oui	oui													
	déchèteries (inspire annexe III)	oui	oui													
Nettoyement		Communes		Grand public												
	tranchées forains	oui	oui													
	poubelles de rue	oui	oui													
	bois de la DP	oui	non													
	toilettes publiques	oui	oui													
	parcours de fauchage	oui	non													
	espaces canins	oui	non													
	espaces conventionnés (CQP à ferme)	oui	non													
VH		Communes		Grand public												
	circuits de VH	oui	oui													
	coffres à sel	non	non													
	dépôts de sel	oui	non													

Illustration 18 : Tableau de priorisation de l'ouverture des données

Le travail de recensement et de priorisation de données à diffuser est primordial pour les porteurs du projet *open data* puisqu'il leur permet d'obtenir un état exhaustif des données géographiques disponibles au sein de l'institution. En outre, chacun de ces administrateurs constitue un relais dans les directions et services permettant de sensibiliser à l'ouverture des données. A l'issue de cette première étape, la définition des données se restreint : sont concernées par la diffusion les données géographiques, en particulier celles soumises à la directive INSPIRE.

L'identité des données diffusables est maintenant beaucoup plus précises : on connaît leur producteur, leur caractéristique et on a même une indication sur leur potentiel de mise à disposition. Cependant, on dispose de cette carte d'identité uniquement pour les données géographiques, et exclusivement celles de la communauté urbaine. Or, ces deux limites

¹ Source : Document interne (Février 2012)

restreignent considérablement l'identification du patrimoine informationnel métropolitain. Pour identifier les données non géographiques et toutes celles produites par des acteurs externes à la collectivité, un deuxième outil est utilisé: la liste des données d'un secteur d'action publique.

3) Lister les données d'un secteur d'action publique

La deuxième étape de l'exploration de l'espace informationnel de l'agglomération lyonnaise consiste à circonscrire un secteur d'action publique et à en recenser l'ensemble des données associées. L'identification des données est ici liée à un usage potentiel de la donnée, résultat de la phase de problématisation de la politique d'*open data* (Chapitre 2). La direction de la voirie souhaite mettre en place une Centrale de mobilité qui rassemble l'ensemble des informations relatives à la mobilité individuelle. Elle va alors lister l'ensemble des modes de transport de l'agglomération, leurs opérateurs et les données pertinentes pour l'information des usagers.

No.	Catégorie de données	Description
1	Données routières : Etats de trafic temps réel produits par le PC Criter	Produite par le Grand Lyon en interne, cette donnée est aujourd'hui déjà mise à disposition d'opérateurs privés conventionnés à cet effet.
2	Données routières : Etats de trafic temps réel historisés	Il s'agit de la même donnée que précédemment, mais archivée par le Grand Lyon.
3	Base créée dans le WP3 d'Optimod pour la mise en correspondance des différentes sources de données	Il s'agit d'une donnée statique, produite par un prestataire pour le compte du Grand Lyon qui s'en assure la propriété.
4	Données parking VP : Localisation et disponibilité en temps réel des parcs LPA	Cette donnée est produite par Lyon Parc Auto, opérateur de droit privé contrôlé majoritairement par le Grand Lyon, la Ville de Lyon et le Conseil Général.
5	Vélo'v (Localisation et disponibilité temps réel des Stations vélo)	Cette donnée est produite par l'opérateur JCDecaux au titre d'un marché qui lui est confié par le Grand Lyon.
6	Horaires théoriques du réseau TCL	Cette donnée est produite par Keolis, l'exploitant du réseau TCL, titulaire d'une DSP attribuée par l'AOT du réseau TCL, le SYTRAL.
7	Temps réel TCL (e-dylic)	Cette donnée est produite par Keolis, l'exploitant du réseau TCL, titulaire d'une DSP attribuée par l'AOT du réseau TCL, le SYTRAL.
8	Temps réel TER par gare	Cette donnée est produite par la SNCF, exploitant conventionné par l'AOT, la Région Rhône-Alpes.
9	Données autoroutières : événements et états de trafic	Il s'agit des données produites par les concessionnaires et commercialisées par Autoroutes Trafic.

Illustration 19 : Tableau listant les données souhaitées pour l'intégration dans la Centrale de mobilité. Cette liste sera par la suite complétée avec les données d'autres acteurs¹.

Cette méthode d'identification permet de définir beaucoup plus précisément les données produites tant par le Grand Lyon que par des acteurs externes. Les données sont clairement identifiées en vue d'un usage précis : la mise en place de services d'information multimodaux. Il ne reste plus ensuite qu'à prendre contact avec chacun de ces acteurs pour obtenir la mise à disposition de leurs données.

Particulièrement efficace, ce procédé d'identification nécessite toutefois de circonscrire un secteur d'action publique et de définir des usages précis des données. Or, en l'absence d'orientation politique particulière, les acteurs en charge du projet d'ouverture de données ne savent pas quels secteurs prioriser.

Le gisement de données du territoire est énorme. Aujourd'hui, on a attaqué la mobilité, mais sur la santé, il y a des données considérables sur le territoire, tous les sujets en fait, il y a un nombre de données considérable, donc on ne peut pas aller partout ni même tout identifier².

Ces limites se manifestent par le ralentissement des données publiées sur la plateforme du Grand Lyon. En effet, une fois les données répondant à la problématisation initiale publiées, comment déterminer le prochain secteur à prioriser ? Quelles données doit-on mettre à disposition ? Qui doit définir les données à publier (les élus ? l'administration ? les utilisateurs ?) ? En fonction de quels critères ? Pour obtenir des réponses à ces questions, les chefs de projet rencontrent des utilisateurs potentiels afin qu'ils leur indiquent des scénarios d'usage et les données associées.

4) Demander aux réutilisateurs

Les porteurs du projet vont faire appel aux réutilisateurs potentiels afin qu'ils leur indiquent les données qu'ils souhaiteraient obtenir. Un des chefs de projet rencontre ainsi les membres de l'association OpenData69. Le collectif lui fournit un document listant des exemples de données (déplacements, développement économique, cadre de vie, environnement, état civil, etc.). De même, la communauté urbaine reçoit régulièrement des emails d'usagers lui réclamant la mise à disposition de telle ou telle donnée. Ces demandes sont également le fait d'entreprises rencontrées de manière informelle lors de salons ou

¹ Source : Document de travail – Projet Optimod (Septembre 2012)

² Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

d'événements divers. Les agents de l'administration sont régulièrement sollicités par des entreprises souhaitant avoir accès à leurs données. Ces demandes leur permettent de constater l'intérêt que suscitent leurs données et justifient leur mise à disposition.

Romain : Il y a beaucoup de boîtes qui nous demandent les données 3D. Donc, on imagine, qu'il y a quand même un réel intérêt. [...]

Enquêteur : Les demandes que vous avez pour les données 3D, c'est en mode, on sait que vous en produisez, quand est-ce que vous les ouvrez ?

Romain : Tout à fait.

Enquêteur : C'était qui ? Des petites boîtes ?

Romain : Plutôt des grosses, souvent des grosses. Veolia, Dassault, Générale des Eaux... Parce que eux, ils sont organisés et ils connaissent. Ils savent. Mais pour autant, les petites sociétés sont également intéressées, clairement. Mais localement. Ils n'ont pas déjà imaginé des applications, alors que les grosses boîtes derrière, elles ont déjà des idées précises¹.

Ces échanges avec les utilisateurs légitiment la politique d'ouverture des données en justifiant de sa pertinence pour des acteurs externes. Ils facilitent également le processus d'ouverture en arguant auprès des producteurs d'une demande pour leurs données, et dans certains cas, en leur présentant des scénarios d'utilisation. Plus que d'identifier des données au sein de l'administration, ils permettent ainsi de prioriser les secteurs d'action publique et les données à ouvrir en fonction de la demande des utilisateurs. Les données sont maintenant définies en fonction du potentiel de réutilisation qu'elles suscitent².

5) Multiplier les réunions exploratoires avec des acteurs divers

L'identification des données passe également par des réunions avec des acteurs pressentis comme de potentiels producteurs de données. C'est en fonction des connaissances individuelles ou au hasard de rencontre ou de lectures que les données et leurs producteurs sont identifiés. Les contacts informels sont une source potentielle de découverte d'un gisement de données. « *Tiens Antoine, toi qui connaît Johann de la mission énergie, penses-tu qu'il aurait des données à mettre à disposition ?* ». « *J'ai lu le rapport de la direction de la propreté, qui indique les tonnages par type de déchets, ça pourrait être une donnée intéressante* ». « *Vous devriez aller voir l'agence d'urbanisme, ils produisent énormément de*

¹ Source : Entretien Romain, producteur de données géographiques (21 novembre 2013)

² Dans une perspective similaire, aux Etats-Unis, la loi pour la liberté d'information (*Freedom of Information Law*) oblige les administrations publiques à transmettre leurs documents à quiconque en fait la demande. Les responsables du projet *open data* de la ville de New York s'appuient sur ses demandes massives pour identifier les données à ouvrir et prioriser leur publication.

données. » Le processus d'exploration s'inscrit dans une perspective de sérendipité, de découvertes heureuses au fil des rencontres.

Ces différentes pistes sont explorées par l'organisation de réunions avec des acteurs de ces institutions diverses. Cependant, bien souvent, ces acteurs n'ont pas l'habitude de traiter de la « donnée ». La donnée n'est en effet que rarement appréhendée en tant que telle. Elle est perçue comme une information, un outil de travail, un tableur Excel, un budget, etc. Une première réaction commune régulièrement observée est : « *Mais, je n'ai pas de données* » ou « *A quoi vous pensez comme données ? Qu'est-ce que vous voulez que l'on mette à disposition ?* ». La notion de données ne va ainsi pas de soi. L'identification consiste également à requalifier une information, un fichier informatique, un document, une couche SIG comme donnée. Les données sont catégorisées comme telles au cours du processus de mise à disposition.

Les porteurs du projet sont alors dans une situation paradoxale : ils rencontrent ces acteurs pour identifier des données, mais ces derniers leur demandent de leur indiquer quelles sont ces données. « *Ils sont marrants, mais comment est-on censé connaître les données qu'a la DLB ?* »¹. Pour résoudre cette contradiction, les acteurs s'appuient sur des données identifiées sur d'autres portails *open data*.

6) Regarder ce qui se fait « ailleurs »

La quatrième étape de notre exploration nous emmène hors du territoire du Grand Lyon pour observer les données publiées sur les portails *open data* des collectivités françaises et étrangères. Le postulat repose sur l'idée que si ces institutions possèdent ces données, la communauté urbaine de Lyon a de grandes chances de les posséder.

Suite à cette discussion, Clément m'a demandé de regarder quels étaient les jeux de données disponibles sur les sites *open data* des autres collectivités territoriales. Cela permet d'une part de dire aux directions « *regarder eux ils l'ont fait, donc on peut le faire aussi* », mais également de savoir quelles sont les données que l'on peut demander aux interlocuteurs définis dans chaque direction. Aujourd'hui, on a qu'une connaissance très partielle des données dont ils disposent. Cependant, Clément m'indique : « *On va faire autant que les autres, mais pas plus que les autres* » (au niveau des jeux de données mis à disposition). Il y a donc une volonté de suivre la tendance de l'*open data*, mais aucunement d'être à la pointe sur ce sujet².

¹ Source : Journal de terrain, Clément (16 avril 2013)

² Source : Journal de terrain (23 janvier 2013)

Cet effet de mimétisme vise également à s'assurer que la communauté urbaine de Lyon ne soit pas à la traîne sur le sujet de l'open data. « Quelles sont les données qui sont mises à disposition par les autres collectivités ? » est une question récurrente des élus et des directeurs de la collectivité. Un travail de benchmarking listant les jeux de données publiés sur les portails open data des collectivités françaises et étrangères est particulièrement utile lors des rencontres avec des acteurs identifiés comme de potentiels producteurs de données.

Le représentant de la ville de Lyon nous a ensuite demandé quelles données on souhaiterait que la ville de Lyon mette à disposition prioritairement. On leur a indiqué que cela dépendait de la stratégie politique (dev éco ou transparence). Problème : la ville de Lyon n'a pas de stratégie de diffusion définie. Néanmoins, plusieurs données complémentaires à celles du Grand Lyon pourraient être mises à disposition, notamment dans les domaines de l'espace public, du handicap et de l'accessibilité, de l'état civil, de l'éclairage et des espaces verts. On leur a fourni une liste de données communales publiées sur d'autres portails open data¹.

¹ Source : Journal de terrain – Réunion avec la ville de Lyon (20 mars 2013)

VIE INSTITUTIONNELLE
Liste des élus (fonction, commission, groupe politique)
Budget primitif
Organigramme des services
Localisation et horaires d'ouverture des services et équipements municipaux
Marchés publics
ETAT CIVIL
Naissance, mariage, décès (stats)
Liste des prénoms
ELECTIONS
Résultats électoraux
Carte électorale
Bureaux de vote
Electeurs inscrits
ORDRE PUBLIC
Vidéosurveillance
VOIRIE
Stationnement (limite de durée, gratuits/payants (prix), nbre de places, places handicapés, cars)
Eclairage public
Espaces verts
Cimetières
Toilettes publiques
URBANISME
Permis de construire
SANITAIRE ET SOCIAL
Crèches
Foyers de personnes âgées
Emplacements des défibrillateurs
ENSEIGNEMENT
Etablissements scolaires
Effectifs
Secteurs scolaires
CULTURE
Bibliothèques
Musées
Ecoles de musique
Salles de spectacles
Manifestations culturelles
SPORTIF ET LOISIRS
Liste des associations
Equipements sportifs
Manifestations sportives
Aménagements touristiques
Aires de jeux pour enfants

Illustration 20 : Liste des données communales publiées sur les sites open data de collectivités¹

*

* *

L'exploration de l'espace des données de l'agglomération lyonnaise repose sur une démarche pragmatique et opportuniste. Elle ne repose sur aucune stratégie déterminée préalablement. En l'absence de cadre planificateur, l'identification du patrimoine informationnel est loin d'être exhaustif. Seule une minorité de donnée est identifiée à l'issue

¹ Source : Document de travail interne (mars 2013)

de cette épreuve. L'épreuve d'identification peut ainsi marquer la fin du processus d'ouverture pour certaines données. C'est le cas de celles dont les porteurs de projet n'ont pas connaissance, ou de celles identifiées, mais qui ne sont pas associées à la plateforme par manque de volonté, oubli ou d'autres priorités.

On est hyper dépendant du réseau que l'on peut avoir en interne. Parce qu'on connaît bien Matthias, on se parle au téléphone, il est de bonne volonté, machin, on le fait. Mais je connais moins Hughes, donc je vais moins le chercher. Et pourtant il est dans le même réseau d'administrateurs, alors du coup, je ne te parle même pas de la nana qui est au fin fond du service truc que je ne connais pas du tout et qui a même jamais entendu parler de notre truc. Donc, du coup, là dedans, on a un biais, un vrai biais de mise à disposition de données¹.

Les limites de l'identification reposent en premier lieu sur la taille du réseau interpersonnel du chef de projet. S'il ne connaît pas certains producteurs, ces derniers ne seront pas associés à l'ouverture des données. En outre, les acteurs invités à publier leurs données sur la plateforme sont essentiellement les partenaires traditionnels de l'action publique communautaire : agence d'urbanisme, aéroport, entreprise en contrat de délégation de service public, municipalités, conseils généraux, etc. Les données produites par des organisations citoyennes sont exclues a priori de la démarche du Grand Lyon.

Jusqu'où doit-on aller dans l'ouverture de notre plateforme ? Est-ce que demain on doit mettre à disposition la liste des caméras recensées par [le collectif militant] Lyon sous surveillance ? On ne le fera pas et pourtant leurs données sont de qualité. Est-ce que l'on doit mettre les données d'Open Street Map ? Les données INSEE, je veux bien, mais les données OSM, je suis plus réservé².

En dissociant ces producteurs, les chefs de projet accordent une visibilité et une légitimité aux données produites par des institutions officielles. Dès lors, à l'issue de l'épreuve d'identification, certaines données sont écartées de la démarche d'ouverture de la communauté urbaine. Seule une minorité de données accéderont à l'épreuve suivante de la publicisation.

II. Publiciser

Etape phare de l'ouverture des données, la publicisation est l'association des données à un nouvel usage et un nouveau public d'utilisateurs. De la même manière que les ingénieurs

¹ Source : Journal de terrain, Réunion Gouvernance de la donnée, (14 octobre 2013)

² Source : Journal de terrain, Clément (Février 2013)

mettent en scène des utilisateurs tout au long de la phase de conception des objets techniques¹, les producteurs préfigurent des usages potentiels à partir desquels ils jugent de l'opportunité de mettre à disposition leurs données. Ils avancent un certain nombre d'hypothèses sur les éléments qui composent le monde dans lequel la donnée ouverte doit prendre place. Ils élaborent des scénarios – ou scripts – mettant en scène des acteurs et l'espace dans lequel ils vont évoluer. Ces représentations varient en fonction des producteurs et des données. Selon une direction, la mise à disposition des données peut mener à une modification des rapports de pouvoir, à une remise en cause de l'action de la collectivité, ou encore à des usages malveillants pour les citoyens. D'autres producteurs de données peuvent être davantage attentifs à des aspects économiques (crainte d'espionnage industriel), juridiques (responsabilité du producteur engagée en cas de mauvais usage), ou sécuritaire (risque pour la sécurité publique et l'intérêt général). Enfin, pour certaines données spécifiques aux processus métiers internes à la collectivité, les producteurs n'imaginent aucun usage potentiel par des acteurs externes et ne voient donc pas l'intérêt de les mettre à disposition.

Afin d'intéresser les producteurs à l'open data, l'enjeu est alors de pondérer ces risques perçus par les avantages potentiels à l'ouverture de leurs données. La donnée est considérée comme publiable quand le producteur estime que les avantages à associer la donnée à de nouveaux utilisateurs sont plus grands que les risques inhérents. La publicisation de la donnée consiste ainsi à représenter un « public fantôme² » et à le traduire en une multitude d' « êtres intermédiaires³ », dont on ne sait s'ils seront les utilisateurs réels des données mises à disposition, mais que l'on mobilise pour déterminer le caractère de diffusibilité des données.

A. Les « porte-paroles » de l'ouverture : quand les producteurs questionnent leurs données par rapport à des usages imaginés

Le travail de sélection des données publiables est généralement confié aux producteurs et non aux chefs de projet *open data*. Les producteurs, dont on juge qu'ils ont la connaissance

¹ Madeleine Akrich, « Les objets techniques et leurs utilisateurs, de la conception à l'action. » dans Bernard Conein, Nicolas Dodier et Laurent Thévenot (eds.), *Les objets dans l'action*, Raisons Pratiques., Paris, Editions de l'EHESS, 1993, p. 35-57.

² Walter Lippmann, *The phantom public*, New York, Simon & Schuster, 1927.

³ Dominique Boullier, « Le client du poste téléphonique : archéologie des êtres intermédiaires » dans *Débordements. Mélanges pour Michel Callon.*, Paris, Presses de l'Ecole des Mines, 2010, p. 41-61.

et le pouvoir nécessaire pour mener la réflexion sur l'ouverture des données au sein de leur organisation¹, deviennent les « porte-parole » du projet d'ouverture des données. Ils sont chargés d'identifier les données de leur organisation, de les questionner et de faire valider leur mise à disposition². Le cas de l'administratrice des données géographiques de la direction de la propreté illustre le processus typique du travail effectué par ces porte-parole.

On nous a dit : « Il y a une plateforme de diffusion qui va s'ouvrir, la première question, c'est de faire un tour de toutes tes données ». Donc, tu dis « voilà, j'ai toutes ces données là ». Ensuite, je suis allé en comité de direction, et je leur ai posé la question de savoir s'ils voulaient publier ces données sur SmartData. Et donc, ils ont dit : « telle donnée : personne, telle donnée : oui, telle donnée : non, etc. » [...] Et puis, le comité de direction m'a demandé de regarder les risques et les avantages à diffuser la donnée. Donc, j'ai pris chacune des données en disant voilà, y'a des risques, des avantages, etc³.

Pour chacune des données recensées, elle entame ensuite un processus de réflexivité sur chaque donnée afin de déterminer sa diffusibilité⁴. Elle identifie des risques et des avantages généraux à la diffusion des données, puis elle interroge chacune des données selon trois critères : technique, environnemental et social.

Risques	Avantages
Changement des rapports de force entre le citoyen et l'administration quant aux décisions à prendre (évaluation, légitimité, efficacité)	Transparence de l'action publique Image moderne de la collectivité
Perte de pouvoir sur la décision	Sens redonné à la relation administration / citoyen
Production de sites discriminants, intrusifs ou source d'inquiétude pour les citoyens	Possibilité de retraitement des données publiques pour la production de la culture, du développement économiques, de nouveaux services, de la prospective, du débat public, de la connaissance...
Manque de fiabilité : discrédite l'action publique et remet en cause le travail de l'administration	Coproduction des mises à jour

¹ Cela n'est pas toujours le cas, certains interlocuteurs se sont révélés être des « porte-paroles » ne représentant que peu leur organisation.

² L'implication de ces porte-parole est très variable. Elle dépend de leur motivation, de leur intérêt personnel pour le sujet, des relations personnelles qu'ils peuvent entretenir avec les porteurs du projet, etc.

³ Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

⁴ L'étendue de ce questionnement ne fait l'objet d'aucune codification et varie fortement d'une direction à l'autre.

Illustration 21 : Document interne de présentation de la démarche de diffusion des données au comité de direction de la direction de la propreté (4 avril 2012)

Pour chacune des données, ce « porte-parole » quantifie ensuite les risques et les avantages à leur mise à disposition selon des critères totalement subjectifs.

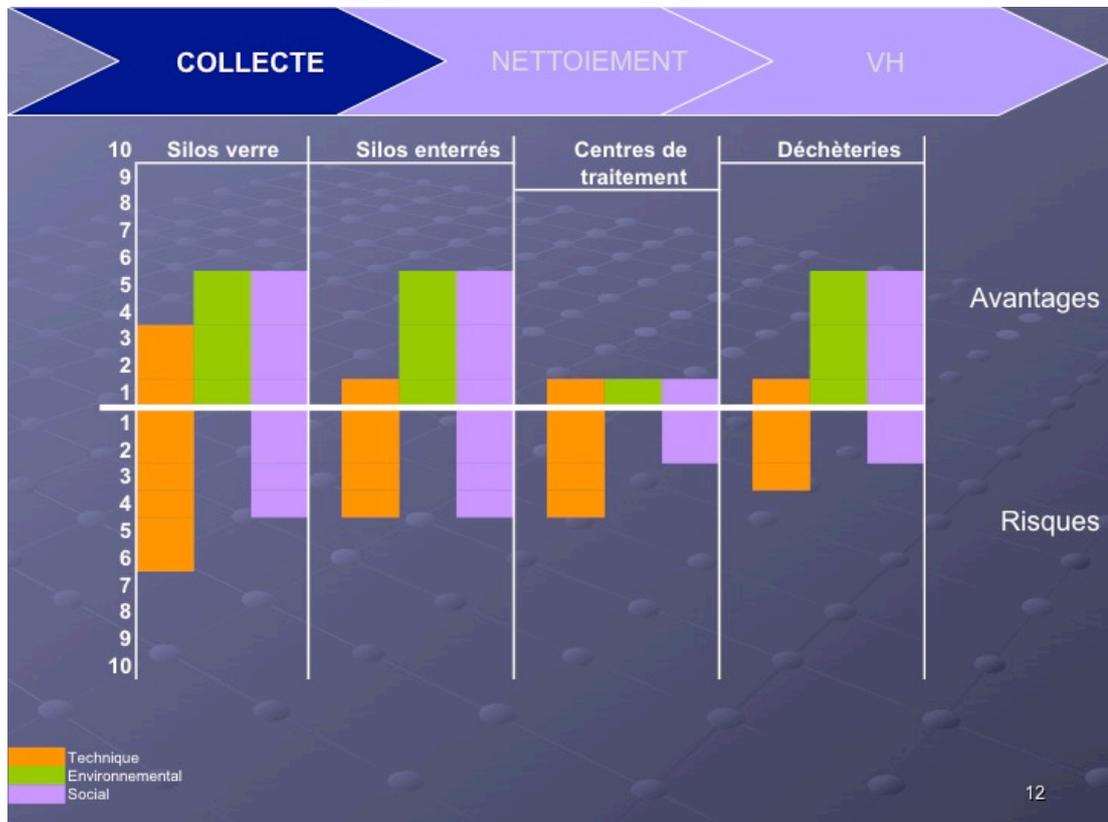


Illustration 22 : Document interne de présentation de la démarche de diffusion des données au comité de direction de la direction de la propreté (4 avril 2012)

Le détail de ce questionnaire appliqué à une donnée particulière, les silos à verre, est le suivant :

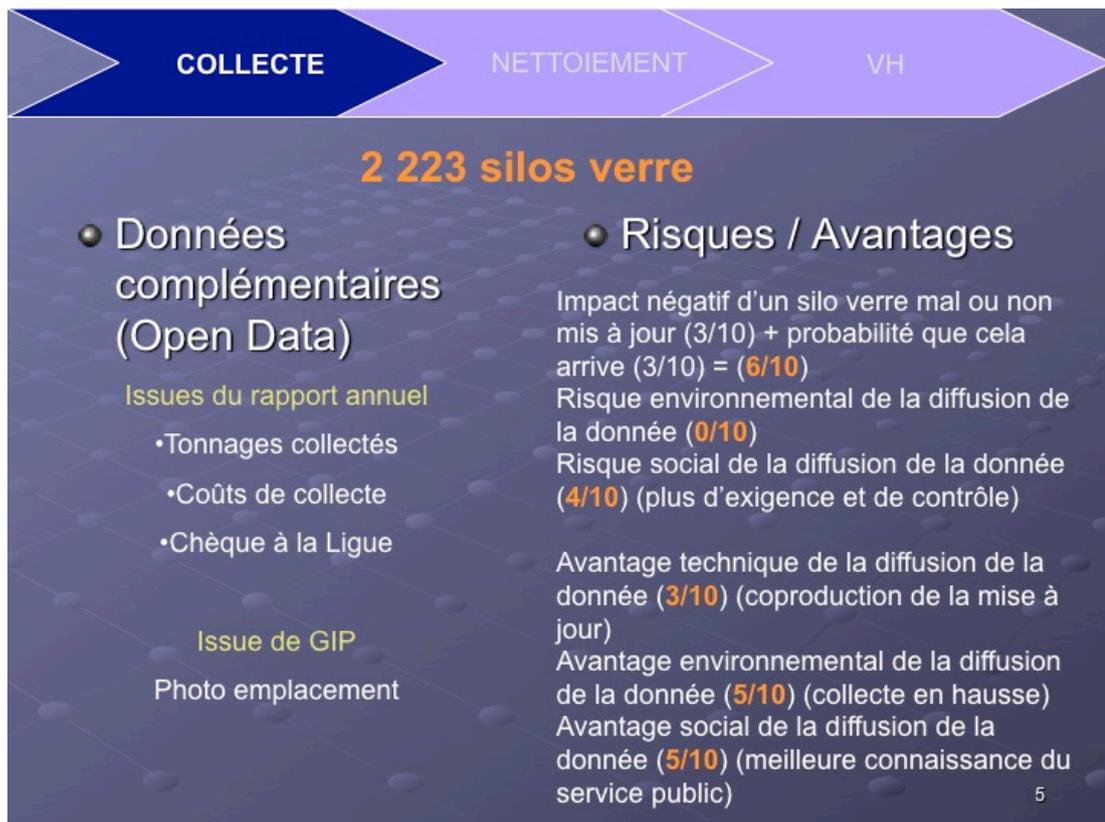


Illustration 23 : Document interne de présentation de la démarche de diffusion des données au comité de direction de la direction de la propreté (4 avril 2012)

L'exemple de la direction de la propreté révèle la pluralité des questionnements sur la donnée. Différents registres de qualification (juridique, technique, environnemental, social, etc.) sont employés pour définir chacune des données. Si la formalisation de ce questionnement sur la donnée est une exception au sein de la communauté urbaine, le processus de réflexion sur la donnée est similaire pour l'ensemble des producteurs de données. La donnée est questionnée selon des critères multiples et les réponses à ces questions sont loin d'être univoques. En outre, celles-ci peuvent évoluer au fil du processus d'ouverture. La même donnée peut alors prendre des identités variables en fonction des caractéristiques dont les acteurs la dotent ou de la situation dans laquelle elle est projetée.

B. Questionner la donnée : associer la donnée à un nouvel usage

Après avoir présenté succinctement la démarche de détermination de la diffusibilité d'une donnée, il est nécessaire de présenter maintenant le détail des registres de questionnement de la donnée. En soulevant des enjeux de propriété, de qualité, de sensibilité,

de responsabilité, de coût et d'intérêt, les producteurs caractérisent précisément leurs données tout en les associant à de multiples utilisateurs imaginés.

1) Propriété et responsabilité

La propriété de la donnée est une question centrale pour déterminer du caractère diffusible ou non de la donnée. La règle en la matière est que le Grand Lyon ne diffuse que les données dont il est propriétaire.

Les données qui ne sont pas à nous, il n'y a pas de question à se poser : elles ne sont pas à nous, je n'ai aucune légitimité pour les diffuser. Je prends l'exemple du cadastre, quand on leur a dit, on va mettre le cadastre, mais juste en version image, pas téléchargeable, la DGFIP [Direction générale des finances publiques] nous a dit, « *non, non, non mettez tout si vous voulez, y'a pas de problème* ». J'ai dit non, je n'ai pas envie. Ce n'est pas à moi, je ne gère pas, je n'ai pas envie de récupérer des questions sur les mises à jour ou autres, ils se débrouilleront. Pareil pour tout ce que j'appelle les données externes, que soit j'achète, soit je récupère au travers de conventions d'échange. Carte géologique, on l'achète au BRGM, pas le droit de diffuser. Corine Land Cover, c'est déjà sur internet. Ce n'est pas à moi, je ne vais pas les diffuser. Le SYTRAL, vaste problème. Les réseaux EDF/GDF, ce n'est pas moi qui vais m'amuser à les mettre sur SmartData¹.

On peut cependant se poser la question lors des nombreux cas où des agents du Grand Lyon croisent des données externes avec des données internes. Est-ce que cette donnée enrichie est une donnée appartenant au Grand Lyon ou reste-t-elle la propriété du producteur initial ? Par exemple, à partir des données fournies par l'INSEE, la communauté urbaine de Lyon cartographie les IRIS (Îlots regroupés pour l'information statistique) en ajoutant pour chacun d'entre eux quelques informations provenant du recensement. En croisant ces données et en les cartographiant à l'IRIS, le service Igéo enrichit la donnée initiale et apporte une plus-value. Se pose alors la question du droit de la communauté urbaine à diffuser cette donnée enrichie.

Peut-être qu'on peut la diffuser. Mais, là, tu vois, moi ça m'interpelle. Finalement, cette donnée, c'est des données de l'INSEE, ce que moi je mets dedans, population, nombre de logements, etc., ça vient directement de chez eux ; les IRIS, c'est quand même eux qui les déterminent, même si c'est moi qui les ait dessinés. Est-ce que je suis vraiment légitime pour mettre ça sur le site SmartData ? Est-ce que l'INSEE ne pourrait pas dire : « *ce sont mes données, je ne veux pas que vous les mettiez* » ? Je me fais peut-être des idées. [...] Sur ce type de questionnements, je ne saurais même pas à qui m'adresser. A qui je demande ? A la DAJCP [Direction des Affaires Juridiques] ?²

La propriété des données est un enjeu qui n'admet pas de réponses évidentes. Le droit reste relativement flou sur la notion de propriété des données et la catégorie de données

¹ Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

² Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

publiques. Juridiquement, il n'existe pas de définition de la « donnée publique »¹. La transposition dans le droit français de la directive européenne sur la réutilisation des données publiques (2003) traduit la catégorie de « données publiques » par les termes de « documents administratifs » et « d'informations publiques »². Selon cette loi, le caractère public d'une donnée n'est pas attaché à l'institution la produisant, mais à la finalité de l'action. Si un opérateur privé produit des données dans le cadre d'une mission de service public, celles-ci seront considérées comme publiques³. La notion de propriété des données n'a pas davantage de statut juridique. Seules les bases de données peuvent faire l'objet d'une protection sous les dispositifs soit du droit d'auteur dans les cas où la structure et l'organisation des données est suffisamment originale pour relever d'une œuvre de l'esprit, soit du droit *sui generis*, protégeant les investissements substantiels réalisés par les producteurs.

En outre, les données sont un enjeu récent pour les pouvoirs publics et la question de leur propriété n'a pas toujours été prise en compte dans les contrats publics, notamment de délégation de service public (DSP). Longtemps considéré comme mineur, le sujet des données apparaît rarement dans ces documents. « *On a beaucoup de contrats qui ont été signés pour du long terme, et à une époque où la question des données n'était pas déterminante*⁴. » Cette absence soulève un flou juridique source de discussions conflictuelles. « *Aujourd'hui, on se bagarre pour récupérer des données des DSP, puisque les entreprises nous disent que le simple fait d'archiver les données leur donne un droit de propriété sur celles-ci et ne veulent pas nous les restituer*⁵ ». La communauté urbaine de Lyon et l'entreprise JC Decaux ont par

¹ La notion de donnée est présente dans la législation française (Articles 226-13 et suivants du code pénal ; Articles L. 341-1 du Code de la propriété intellectuelle (protection des droits des producteurs des bases de données) ; Loi Informatique et Liberté de 1978).

² « Sont considérés comme documents administratifs, au sens des chapitres Ier, III et IV du présent titre, quels que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, directives, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions et décisions. » (Article 1^{er} de la loi n°78-753 du 17 juillet 1978)

³ Néanmoins, si les données sont utilisées dans le cadre d'une mission de service public industriel et commercial, elles ne sont pas considérées a priori comme publiques. En outre, la législation française fixe des exceptions au caractère public d'une donnée : les données relevant de la vie privée, de la sécurité nationale ou sur lesquelles des tiers détiennent des droits de propriété intellectuelle ne sont pas considérées comme publiques.

⁴ Source : Entretien avec Harald, chef de projet, Grand Lyon (8 janvier 2013)

⁵ Source : Journal de terrain (6 février 2013)

exemple négocié longuement pour déterminer la propriété des données relatives au marché public de vélos en libre service¹.

La propriété et le statut public de la donnée ne sont ainsi pas aisément définissables. Face aux nombreux cas de figure, les directions se retrouvent parfois dépourvues comme le souligne cet interlocuteur du service mobilité : « *Cela n'est pas si évident que ça puisse paraître. On a fait bosser un cabinet d'avocats spécialisés en économie numérique pour établir les différents cas possibles* »². Le recours à des expertises externes est alors nécessaire pour clarifier cet enjeu déterminant la diffusibilité : une donnée peut ne pas être diffusée à cause d'un manque de clarté sur celui qui a les droits ou la responsabilité pour le faire³.

Complexe lorsque la chaîne de la donnée s'allonge, la question de la propriété est indissociable de celle de la responsabilité. Est-ce que la propriété de la donnée instaure un devoir de responsabilité vis-à-vis de celle-ci ? La réponse à cette question est complexe. Dans les usages internes, le producteur de la donnée est souvent tenu pour responsable en cas de mauvaise qualité de sa donnée : on peut lui demander des comptes et il peut être sanctionné⁴. Qu'en est-il lorsque cette donnée est diffusée à des utilisateurs externes ? Cette interrogation est l'une des craintes récurrentes soulevées par les producteurs.

Le représentant de la ville de Lyon s'inquiète de la responsabilité de la ville de Lyon en cas de données de mauvaise qualité. En effet, deux cas soulèvent aujourd'hui cette question au sein de la ville, qui a été assignée au tribunal administratif parce que l'horaire du stationnement gratuit qu'elle indiquait sur son site internet ainsi que les voies interdites à la circulation en vélo au sein du parc de la Tête d'or, étaient erronées. Des usagers ont eu une amende, et se reportent aujourd'hui contre la ville. Est-ce qu'il n'y a pas un risque avec les données ouvertes ?⁵

¹ « *Juridiquement, il faut déjà voir à qui appartiennent les données Velov. Moi, je pense que c'est au Grand Lyon parce qu'il s'agit d'un marché public, non d'une DSP. Mais Decaux n'est pas tout à fait d'accord avec cela. Il faut donc qu'on soit clair là dessus. Si les données nous appartiennent, on est libre de les mettre à disposition.* » (Harald, journal de terrain, 6 février 2013). Ces négociations se sont résolues par la mise à disposition en open data des données de VLS par l'entreprise JC Decaux.

² Source : Journal de terrain (6 février 2013)

³ Ces préoccupations sur la propriété ne sont pas propres à l'administration : elles sont similaires dans le milieu académique comme le pointe Christine Borgman dans son étude sur la diffusion des données de la recherche. Christine L. Borgman, *Big Data, Little Data, No Data: Scholarship in the Networked World*, Cambridge Mass., MIT Press, 2015, p. 218.

⁴ La responsabilité peut être définie comme « *la qualité de ceux qui doivent, l'irresponsabilité la qualité de ceux qui ne doivent pas, en vertu d'une règle, être choisis comme sujets passifs d'une sanction.* »

Paul Fauconnet, *La responsabilité : étude de sociologie*, Paris, Félix Alcan, 1920, p. 11.

⁵ Source : Journal de terrain

A partir du moment, où l'on fait SmartData, se pose la question de la responsabilité du service public. Je vais prendre un exemple très concret sur la donnée de gaz par exemple. Quelqu'un peut gérer la localisation des réseaux par rapport à son usage qui est propre, à partir du moment où il publie la donnée, et où quelqu'un va faire des travaux en s'appuyant sur cette donnée là. Et s'il y a des accidents, des morts, etc. Quelle est la responsabilité des uns et des autres par rapport à ça ? Quelle responsabilité résiduelle on peut imputer au producteur de données sur un usage que lui ne connaît pas a priori ? Sachant que lui, il n'a jamais intellectualisé qu'il pouvait être responsable de quoi que ce soit, sachant que pour lui, la donnée est un outil¹.

Ces interlocuteurs s'interrogent sur le transfert de responsabilité : est ce que la diffusion des données entraîne une déresponsabilisation du producteur ? Dans quelle mesure peut-il être l'objet d'une sanction en cas d'usage de ses données ? Les acteurs n'ont souvent pas de réponses claires et précises à ces interrogations complexes qu'ils soulèvent et qui déterminent pourtant la diffusibilité des données.

2) Qualité

Peut-on diffuser une donnée qui n'est pas de bonne qualité ? Cette question est récurrente et interroge ce qu'est la « qualité » d'une donnée. Le cas de la donnée « silos à verre » est particulièrement révélateur de cet enjeu. Produite par la direction de la propreté de la communauté urbaine, cette donnée géographique a été successivement publiée, dépubliée, puis republiée sur le site *open data* en l'espace de quelques mois. Ces hésitations soulignent les conceptions variables de la « qualité » d'une donnée selon les acteurs.

Mise à disposition, la donnée est en effet exposée à des regards externes. La crainte de la direction de la propreté était que ces utilisateurs mettent en question le statut de « réalité » de la donnée, et, révèlent in extenso, la méconnaissance du Grand Lyon de son territoire.

Et, après voilà, il y avait l'argument de la qualité de la donnée : on diffuse de la donnée mais elle est pas juste. [...] C'était surtout ça en fait qui les embêtait [la direction] : de livrer des données pas à jour².

En gros, ils se disent que les maires pourraient très bien aller télécharger la donnée et dire que les silos ne sont pas là, qu'ils sont ailleurs. En fait, il y a un peu une crainte de se dire « *on va voir nos données, et si elles sont pas à jour, c'est pas bien.* » Donc, ils voulaient pas les mettre parce que les silos ça bouge souvent et du coup, c'est pas simple à mettre à jour, et donc ils préféreraient ne pas mettre les silos plutôt que de les mettre et d'avoir des réclamations derrière en disant « *vous dites qu'il y a un silo là, alors qu'il n'y en a pas* »³.

¹ Source : Journal de terrain, Tom, direction des systèmes d'information (19 décembre 2013)

² Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

³ Source : Entretien, Louis, administrateur de données, Grand Lyon (22 novembre 2013)

Après avoir été publiée, la direction va alors faire machine arrière et retirer cette donnée le temps de s'assurer de la fiabilité des informations. La donnée va alors faire l'objet d'une campagne de mise à jour et de vérification par des relevés terrain de l'emplacement des silos à verre.

[Le producteur de la donnée silos à verre] m'a dit, en décembre, après le marché de collecte, une fois que tout se sera calmé, on va relancer une campagne de qualité sur les silos. Donc je pense qu'il a du envoyer des gens pour vérifier avec un plan, etc. Donc, il s'est dit après ça, on les met en ligne¹.

[Le comité de direction] m'a demandé de relancer en montrant qu'on mettait bien à jour la donnée. On venait juste d'avoir une très grosse campagne de mise à jour des données de saisie. J'ai sorti plusieurs statistiques sur plusieurs mois comme quoi on mettait bien à jour la donnée, et du coup, on a eu l'aval pour intégrer les données dans SmartData².

L'ouverture des données est un moment privilégié de réflexivité sur les données et leur fiabilité. Cependant, la question de la qualité d'une « donnée » n'a pas de réponse univoque. Les débats sur la qualité de la donnée mettent en scène des cultures différentes de la donnée selon les acteurs qui interviennent. Alors que la direction générale de la Direction de la propreté refuse de publier une donnée qui n'est pas parfaitement représentative et exhaustive, les producteurs de données rétorquent qu'une donnée ne peut jamais être fiable à 100%. A contrario, ils mettent en avant la solidité de la chaîne de production de la donnée pour questionner l'argument de la mauvaise qualité de la donnée « silo à verre » avancé par la direction.

Les retours des personnes sur le terrain, c'était, 80/20. On était dans la même moyenne que le reste des données : à 80% les données sont justes. Y'a des données où tu te dis, y'a pas de suivi. Mais là, c'était pas le cas. Il y avait vraiment des fiches de suivi qui étaient faites, t'as toute une unité, une unité silo qui est créée, qui est en relation avec les subdivisions, il y a pas mal de maintenance sur les silos. Donc, tu vas les mettre dans des magasins pour les réparer, les remplacer, etc. Et avec ce type là qui s'occupait de cette unité et les subdivisions, y'avait vraiment un échange constant sur les silos. Si ça n'avait pas été le cas, tu te dis effectivement, cette donnée là... Mais là, c'était bien suivi, bien cadré, donc, on était dans les bons en terme de qualité.

L'argument de la mauvaise qualité d'une donnée pour refuser son ouverture est ainsi davantage avancé par le comité de direction de la direction de la propreté que par les producteurs qui sont conscients du caractère construit de leurs données. Les entretiens avec les administrateurs de données d'autres directions confirment cela.

¹ Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

² Source : Entretien, Louis, administrateur de données, Grand Lyon (22 novembre 2013)

Est-ce que la qualité a été un critère ? Non. Nous du moment que la donnée est exhaustive, qu'elle a un intérêt... J'ai presque envie de dire, parfaite ou pas, nous, c'est des données qu'on utilise, donc on estime qu'elles sont utilisables. Après, elles ne sont jamais parfaites, mais on les met en ligne¹.

En mettant en avant l'usage quotidien de leurs données, ces acteurs réfutent ainsi l'argument de la mauvaise qualité. Celle-ci ne peut se percevoir que par rapport à l'usage qui est fait de la donnée. D'ailleurs, certains vont même jusqu'à affirmer que la notion de qualité n'a que peu de sens en tant que telle.

Enquêteur : Et grosso modo, tu estimes la qualité des données à... Elles sont fiables à quoi ? 80% ?

Etienne : C'est tellement compliqué. Ça veut pas dire grand chose la qualité des données. [...] La qualité effectivement par rapport à un usage, et puis, il y a tellement de critères de qualité, de manière de la mesurer... Quand on me demande comme ça, de but en blanc, quel est le niveau de qualité des données, je suis embêté pour répondre.

Enquêteur : Mais, parce que c'est une crainte de certains producteurs, de ne pas diffuser leurs données parce qu'ils savent qu'elle n'est pas à jour ou de bonne qualité.

Etienne : Moi, je ne suis pas du tout sur ce créneau là, au contraire. Je suis pour qualifier au maximum la donnée, avec les limites que j'imagine, la qualité pour quoi faire, et sur quels critères exactement on la mesure. Après, je peux te répondre plus en termes d'exhaustivité. Je pense que sur le tracé des voies, on est bon à 95%. Et les noms aussi. Sur les numéros, on est plutôt entre 80 et 90%. [...] En gros, la qualité, il faudrait la mesurer couche par couche et selon des critères qui peuvent être positionnement géographique, exhaustivité, cohérence, etc. Il y a tout un tas de critères [...]. Mais voilà, la qualité pour qui, pour quoi et selon quels critères. Là, tu peux commencer un peu à répondre. Avant... Moi j'ai envie de dire que mes données sont bonnes [rires]. Donner un chiffre, c'est impossible².

En acceptant de publier des données qu'ils savent non parfaitement fiables, les producteurs perçoivent au contraire la mise à disposition de leurs données comme une opportunité d'en améliorer la fiabilité.

J'ai essayé de dire que c'était pas possible d'avoir des données 100% juste, mais au contraire en mettant de la donnée en diffusion, on pouvait avoir le retour des usagers, et nous permettre de la corriger, c'était intéressant aussi³.

L'autre critère, c'est de dire on a pas l'exhaustivité, sur ce, moi j'ai balayé en disant peu importe, on s'en fout. Ça me permet justement d'avoir un retour terrain, si y'a des utilisateurs qui me font remonter via SmartData qu'ils ont trouvé un point d'eau qui n'est pas référencé, ou que le point d'eau référencé, il n'est pas là, mais là. Ça nous permet d'avoir un contrôle terrain, et un retour terrain d'utilisateur. Je suis très pour ce type de retour⁴.

¹ Source : Entretien Martin, administrateur de données, Grand Lyon (6 décembre 2013)

² Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

³ Source : Entretien Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

⁴ Source : Entretien Thibaut, administrateur de données, Grand Lyon (4 décembre 2013)

L'enjeu se déplace alors de la fiabilité initiale de la donnée à la capacité des producteurs à prendre en compte les retours des utilisateurs et à mettre à jour la donnée.

Dans toutes les données, on sait qu'il y a 80% des données qui sont justes et 20% qui sont fausses. Quelle que soit la donnée. La domanialité, elle n'est pas juste. Les voiries, qui sont des données de référence, elles ne sont pas justes. Parce qu'on est une énorme collectivité, et que ça bouge sans arrêt. Tu ne peux pas être à la minute même au fait de la modification. Donc, les silos à verre. Ça bouge les silos. Ils sont enlevés, parce qu'il y a des chantiers, parce que ça fait du bruit, enfin voilà, ça bouge énormément. Donc, t'as pas un agent derrière chaque silo pour te dire, ça y est il vient d'être enlevé, ou celui-là, il bouge pas, donc il faut l'admettre que ce ne soit pas juste à 100%. Mais, moi, je disais ça ne me dérange pas que quelqu'un vienne me dire : « *le silo, la donnée est fausse, ce silo là, il a été enlevé il y a deux semaines, etc.* » Par contre, ce qui me gênerait, c'est qu'à partir du moment où il te le dit, tu ne fasses pas la modification. C'est là qu'il faut être réactif, c'est dans la correction. Par contre, on peut admettre que les données ne soient pas exactes.

En insistant sur la solidité de la chaîne de d'objectivation de la donnée, ces producteurs rejoignent les conclusions d'Alain Desrosières qui affirme que la fiabilité de la donnée « *doit être entendue non pas comme une ressemblance fidèle à la réalité, mais comme une consistance et une robustesse suffisantes pour supporter sans dommage des épreuves variées, transports, combinaisons, critiques des équivalences conventionnelles ou des frontières. Cette « fiabilité » est inextricablement technique et sociale. Elle suppose de longues chaînes d'enregistrements, de calculs et de mises en forme, qui pris ensemble suscitent ou non la confiance de la société* »¹.

3) Sensibilité : la figure de l'utilisateur malveillant

Le caractère sensible d'une donnée est un argument central pour refuser sa diffusion. Pourtant, il n'existe pas de critères objectifs permettant de déterminer la sensibilité d'une donnée. Une donnée n'est jamais sensible par nature. Cette caractéristique ne résulte que de la préfiguration de l'usage qui peut être fait de la donnée. L'embarras de ce producteur pour justifier le choix de sa direction de ne pas diffuser la donnée des circuits de collecte révèle la construction de scénarios d'usage malveillant des données pour caractériser leur sensibilité.

Les circuits de collecte, c'est un peu sensible. Pour quelles raisons ? Je n'arriverai pas à le dire. Je ne sais pas pourquoi ils n'ont pas voulu mettre les circuits de collecte. Ou peut-être se sont-ils dits : « pas tout de suite ». Limite, je trouve ça moins sensible qu'un circuit de neige, parce que dans un circuit de neige, on ne passe pas forcément partout. Y'a des priorités qui sont faites. Alors que dans la collecte, on est sensé collecter tout le monde. Après, y'a peut-être les histoires, de « je

¹Alain DESROSIERES, *Pour une sociologie historique de la quantification*, op. cit. ;

On retrouve ici la conception de la réalité de Bruno Latour : « *Est réel ce qui résiste dans l'épreuve* » Bruno LATOUR. *Pasteur : guerre et paix des microbes : Suivi de Irréductions*. Paris : La Découverte, 1984, p. 244.

« passe trois fois par semaine dans cette commune ou dans ce type d'habitat, je passe qu'une fois dans celui-là »¹.

Dépendante du pouvoir divinatoire des acteurs, les critères de sensibilité d'une donnée peuvent être multiples : sécurité publique, vie privée, raisons politiques, mise en œuvre de l'action publique, image de l'institution, etc. La sensibilité de la donnée fait émerger la figure de l'utilisateur malveillant. Les acteurs imaginent de nombreux cas d'utilisations malveillantes des données basés soit sur leurs expériences antérieures, soit en prêtant des intentions aux utilisateurs. Par exemple, ces acteurs soupçonnent, qu'en cas de diffusion de l'emplacement des coffres à sel ou des locaux d'agents, des utilisations malveillantes qui nuiraient à la bonne mise en œuvre de la politique de propreté.

Les coffres à sel par contre, je suis sûr qu'ils n'y sont pas et qu'ils n'y seront jamais. Parce que les coffres à sel sont destinés aux agents du Grand Lyon pour saler les voiries. Sauf, que le grand public pense que ces coffres sont à destination du grand public, et des personnes iraient prendre du sel. Alors, que ce n'est pas du tout pour le grand public. Et du coup, l'idée c'est pas qu'on leur dise où sont les coffres pour qu'ils aillent se servir².

Par exemple, les locaux d'agents. Y'a à peu près 200 locaux d'agents dans le Grand Lyon, des cantonniers qui balayent, y'a des locaux qui sont dispersés dans les villes, mais tu les vois pas forcément, avec un vestiaire, avec un point d'eau, une table, pour pouvoir se reposer, se mettre à l'abri, ou manger. Et ces locaux d'agents, le risque de diffuser ces points là, c'est que, on vienne déposer devant des sacs de poubelle, en disant il est là, il fera le nécessaire, des piles, de la peinture, enfin que ça devienne une mini-déchetterie³.

De manière similaire, régulièrement, les réticences des acteurs à publier les données portent sur la crainte de polémiques « stériles » qu'elles pourraient susciter, ce qui viendrait entraver le cours des politiques urbaines. En effet, les services seraient alors tenus de répondre à l'utilisation de ces données en proposant une contre-expertise. Une des craintes est notamment que ces informations soient (mal) utilisées et nuisent à l'image de l'institution comme le témoignent ces deux producteurs.

On avait des craintes avant de mettre à disposition nos données, notamment pour l'image de l'entreprise et du service. Le risque que l'on avait pressenti était que des personnes utilisent les données et en fassent une analyse fautive, qui conduise à des polémiques stériles, que l'on doive gérer, mais qui nuisent à l'image, par exemple, quelqu'un qui s'amuse à ne compiler, compter que les stations qui sont toujours vides ou des choses comme ça⁴.

¹ Source : Entretien, Louis, administrateur de données, Grand Lyon (22 novembre 2013)

² Source : Entretien, Louis, administrateur de données, Grand Lyon (22 novembre 2013)

³ Source : Entretien Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

⁴ Source : Entretien Thomas, direction des systèmes d'information, JC Decaux (27 août 2014)

Il y a des gens qui ont dit, oui mais nous, dans nos réunions publiques, pour l'anneau des sciences, il ne faut pas que quelqu'un se mette à sortir : « voilà, j'ai récupéré tout l'historique du trafic », et que ça mette en difficulté des projets, parce que les données, elles peuvent être effectivement utilisées à bon escient par le public, mais elles peuvent être aussi, c'est des données très techniques, pas forcément interprétées comme il faudrait le faire, et du coup, fragiliser certains projets par une utilisation détournée. Tout le monde n'est pas gentil (rires). [...] La crainte de ces personnes étaient que ça soit détourné, mal compris, et mal utilisé alors qu'au départ, nous on va utiliser les mêmes données pour construire notre projet. Enfin, que ça rentre en contradiction avec l'analyse Grand Lyon¹.

Ces exemples soulignent que la donnée peut être considérée comme sensible lorsque sa publication risque de venir perturber la conduite ordinaire de l'action publique. L'argument des compétences techniques nécessaires à l'utilisation des données est également récurrent pour justifier de la crainte d'un mauvais usage.

Une des figures les plus récurrentes d'utilisateur malveillant est celle du terroriste. Par exemple, la crainte de contamination du réseau d'eau justifie le caractère sensible de la donnée du réseau d'eau potable. Régulièrement invoqué, ce critère de « sécurité publique » empêche la diffusion de données. Il n'est pourtant pas substantiel à la donnée comme l'illustre la publication en *open data* de ce même jeu de données par la ville de Vancouver².

Une autre comparaison internationale souligne le caractère relatif de la sensibilité des données. Elle concerne les données à caractère personnel. Si des informations relatives à un individu sont contenues dans une donnée, elles ne peuvent être diffusées. Néanmoins, la question de la donnée personnelle est variable selon les cadres légaux. Une donnée personnelle n'est pas forcément une donnée nominative. La liste des élus est par exemple une donnée nominative, mais ce n'est pas une donnée personnelle car elle ne relève pas de la vie privée. De même, la liste des salaires des agents de la ville de Chicago est une donnée nominative, qui serait considérée comme personnelle en France, mais qui est considérée comme publique par cette municipalité. La transparence s'applique quand les personnes interviennent dans le cadre d'une activité publique. La question du caractère personnel de la donnée n'est pas liée à la donnée en elle-même³, mais à l'environnement dans lequel elle est insérée.

¹ Source : Entretien Martin, administrateur de données, Grand Lyon (6 décembre 2013)

² Source : <http://data.vancouver.ca/datacatalogue/waterNetwork.htm> (Consulté le 28 avril 2016)

³ Le caractère personnel d'une donnée varie également selon le degré de granularité de l'information. Les règles de secret statistique visent à rendre impossible l'identification des individus par l'agrégation des données à une échelle plus large.

La sensibilité varie également dans le temps. Un contexte électoral ou un climat social conflictuel au sein de l'administration, et des données jugées non sensibles vont subitement le devenir.

Sur le passage en comité de direction de la direction de la propreté, Pauline est passée une première fois, ils étaient prêts à tout ouvrir, les données ont été rapidement disponibles sur la plateforme. Puis, elle a effectué un second passage en comité de direction pour leur faire un suivi sur la diffusion des données. Entre les deux, une grève importante a secoué la direction de la propreté. « *Tout le monde était très tendu. Quand je suis repassé en comité de direction, ils avaient peur de tout, ils ne voulaient plus rien ouvrir. Ils avaient peur que les citoyens remarquent que les données ne sont pas à jour ou complètes¹.* » Ils sont alors revenu sur la décision de publier la donnée silos à verre au motif qu'elle n'était pas d'assez bonne qualité. Cette donnée sera publiée à nouveau suite à une campagne de mise à jour.

On a reçu une demande sur SmartData d'accès aux résultats électoraux des municipales de 2001 et 2008 par bureaux de vote. J'ai transmis cette demande à la ville de Lyon. Sans réponse de leur part après deux semaines, je les ai relancés aujourd'hui. Benoît m'appelle dans la foulée, en m'indiquant qu'il a bien reçu la demande initiale, et qu'ils ont eu l'accord pour diffuser cette donnée à la journaliste, mais pas pour la mettre sur la plateforme « en cette période électorale ». Son discours transpirait une paranoïa et une crainte vis-à-vis de la diffusion de ces données : « *C'est une donnée qui est gérée par la préfecture, on aurait pu botter en touche en les renvoyant vers la préfecture. [...] Et vous avez vérifié qui c'était ? A-t-on un moyen de l'identifier ? Parce qu'elle dit être journaliste à Rue89, mais elle envoie depuis une boîte mail perso, ça pourrait être n'importe qui* »².

Enfin, la sensibilité peut également dépendre de la visibilité de l'infrastructure de diffusion et des facilités de réutilisation du format de la donnée. Par exemple, certains interlocuteurs ne sont pas contre une diffusion des données sous un format PDF difficilement retraitable informatiquement, mais ils refusent lorsque sont évoqués la diffusion sous d'autres formats.

J'ai soumis l'idée de publier les données contenues dans ce bilan social sous format brut. Cela lui a posé davantage de réticences. « *Je veux bien mettre à disposition le bilan social, mais la mise à disposition des fichiers sources m'interpelle davantage. On ne va pas leur faciliter la réutilisation pour qu'ils fassent n'importe quoi avec ces chiffres* »³.

Un exemple similaire concerne la liste des marchés publics, qui est publiée, mais dont on ne souhaite pas favoriser la visibilité et encore moins la facilité de retraitement.

Aujourd'hui la liste des marchés publics est disponible quelque part sur le site du Grand Lyon dans un pdf. On est obligé de le publier. Cependant, il y a des réticences à le publier sur le site open data. On ne cherche pas trop à lui donner plus de visibilité... Il est sur le site, mais on le met pas en avant. Il faut bien le chercher...

¹ Source : Entretien Pauline, administratrice de données (14 octobre 2013)

² Source : Journal de terrain (21 janvier 2014)

³ Source : Journal de terrain (7 mars 2013)

Ainsi, une donnée n'est jamais sensible par nature. Le degré de sensibilité d'une donnée évolue dans le temps et en fonction du contexte, des techniques d'analyse ou des valeurs et des cultures professionnelles et nationales. La sensibilité est le résultat d'une épreuve au cours de laquelle les acteurs l'insèrent dans un ensemble de situations particulières. Les producteurs pratiquent un exercice de divination grâce à la multiplication des expériences de pensée par lesquelles ils soumettent la donnée à des situations, des contextes, aux mains d'acteurs qu'ils dotent d'intentions plus ou moins louables. Au travers de ces exemples de risque de mauvaises utilisations des données publiées, les agents définissent, en creux, une norme « morale » sur ce qu'est le « bon » usage des données publiques : l'ensemble des utilisations qui ne vont pas à l'encontre de la collectivité, de ses politiques publiques et qui ne nuisent pas à l'activité commerciale des délégataires de service public, etc. Une hiérarchisation est donc établie entre les « bons » et les « mauvais » usages de la donnée.

4) Intérêt : l'absence d'utilisateurs imaginés

Autre point définissant la diffusibilité de la donnée est son « intérêt » pour des utilisateurs externes au Grand Lyon. Pour définir l'intérêt d'une donnée, les producteurs préfigurent un espace de réutilisation possible en associant leurs données à un nouvel usage et à un nouveau public d'utilisateurs. S'ils n'y parviennent pas, ils ne diffuseront pas leurs données.

Je ne pense pas qu'aujourd'hui il y ait de nouvelles données candidates [à la diffusion]. [...] Les points de collecte [des déchets] ? Je ne sais pas si ça a un intérêt. Après, les cantonnements, c'est plus spécifique métier, c'est vraiment tronçon de désherbage, de fauchage, je ne sais pas si il y a un intérêt¹.

La DLB veut même nous donner la consommation annuelle de carburant du Grand Lyon. Mais bon, est-ce pertinent ? Qu'est-ce qu'on va en faire ?²

Au cours d'une réunion de présentation de la démarche *open data*, une responsable d'une des communes de la communauté urbaine, déclare que « *On n'a pas de volonté d'ouverture. Les données sont utiles pour nos besoins quotidiens, elles ne sont pas intéressantes pour le grand public. Qu'est ce que le Grand public va faire de la donnée sur l'éclairage public ?* »³.

¹ Source : Entretien, Louis, administrateur de données, Grand Lyon (22 novembre 2013)

² Source : Journal de terrain, Clément

³ Source : Journal de terrain

Après, je comprends le collègue de la direction de l'eau, leurs données, à part les terroristes, je ne vois pas trop quel est le genre de public... Si nous, on ne constate pas de choses qui émergent, avec des données qui paraissent utiles au premier abord pour celui qui fait du vélo, etc.¹

Parce que les producteurs ne parviennent pas à se représenter un public d'utilisateurs, certaines données ne sont pas catégorisées comme diffusables.

5) La crainte d'un changement des relations de pouvoir

Enfin, la dernière interrogation déterminant la diffusibilité des données est celle de la transformation des relations sociales que pourrait induire leurs diffusions. A nouveau, les producteurs imaginent des situations qui feraient évoluer, à leur détriment, les relations qui fondent leur pouvoir. Cette crainte est exprimée tant par des acteurs publics que privés, et s'inscrit autant dans des registres organisationnels que politiques ou commerciaux.

Une entreprise en délégation de service public peut craindre que la diffusion de ses données fasse émerger des indicateurs critiques quant au service effectivement fourni. Elle ne va alors pas diffuser l'information permettant la production de ces chiffres alternatifs.

Il y a aussi des choses par exemple qui ne peuvent pas être calculées à partir des données que l'on met à disposition, notamment sur les clauses de niveau de service qui sont insérées dans les contrats. C'est l'ensemble des comptes que l'on doit rendre aux collectivités sur la base de telle ou telle problématique, comme les stations de proximité par exemple. On leur fournit des indicateurs annuels d'évaluation. Si quelqu'un voudrait faire un indicateur alternatif au notre, ce ne serait aujourd'hui pas possible, parce qu'il n'a pas forcément l'information nécessaire sur ce qu'est une station de proximité par exemple, quelle station en est une, quelle station est une station de rotation ou pendulaire, et que l'on ne prend pas en compte dans nos indicateurs. Cette information n'est pas dans les données mises à disposition par exemple. On a été particulièrement attentif à ce point, c'était notre crainte principale avant d'ouvrir : qu'il y ait des mauvaises interprétations des données sur le niveau de qualité de service, avec des indicateurs faux, ou sur des critiques de fonctionnement².

Des acteurs publics redoutent que le développement de nouveaux services à partir de leurs données entraîne une perte de visibilité de leurs missions, qui pourrait se traduire par une perte d'influence dans le jeu institutionnel de la gouvernance urbaine.

Oui, on a un souci de visibilité de chacun des opérateurs. C'est ça le conflit qu'il y a. [...] Les exploitants s'effacent derrière la vision usager par rapport au service qui est offert. Par contre, côté exploitant, moi j'ai envie d'exister, j'investie x millions d'euros tous les ans, c'est pour qu'on

¹ Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

² Source : Entretien Thomas, direction des systèmes d'information, JC Decaux (27 août 2014)

sache que c'est moi qui l'ai fait aussi. [...] Donc, il y a pour eux un enjeu de visibilité et de perte d'influence auprès de l'utilisateur. De notoriété en fait¹.

Etienne : Il y a un piège que je veux éviter, j'ai peur que dans l'esprit des décideurs ou des élus, le fait de voir revenir nos informations au travers d'applications ou de services ou je sais pas quoi estampillés de quelqu'un d'autre qui va dire, je vous ai fait une carte où vous retrouvez les silos à verre, ou je sais pas quoi. Donc ça, ça va arriver sur Smartphone, et à un moment, on a peut-être un élu, qui va nous dire, à quoi ça sert que la voirie continue à saisir les silos de verre, ils sont directement sur mon smartphone. Et qu'ils fassent pas forcément le lien entre le fait que pour réussir à faire l'application sur smartphone, c'est bien les gurgusses du SIG, de la voirie, de la propreté, qui ont bossé, qui ont mis à jour leurs trucs et qu'ils les ont mis dans SmartData, qui font que le service existe. Et pas l'inverse. Parce que ça, ça m'est déjà revenu à l'oreille. A quoi ça sert ? Regarde on est 12 personnes dans mon unité pour saisir de la donnée, alors que sur Google on a tout gratos quand on se connecte sur Internet.

Enquêteur : Du coup, la crainte, c'était que ça vous invisibilise un peu dans l'application finale ?

Etienne : Il y a des chances. Ca il faut pas se leurrer. Il y a de fortes chances. Et je suis pas sûr qu'on soit capable de dire, que les données qu'il y a dans une appli ou qu'il y a dans un service web ou je sais pas quoi viennent de chez nous².

Enfin, les suspicions d'espionnage industriel et de concurrence commerciale restreignent la diffusibilité des données. Cet argument est variable selon la maturité du secteur commercial de l'entreprise comme le prouve la différence de positionnement entre les entreprises Bolloré et JC Decaux.

Après, il y a des acteurs privés, pour lesquels leurs données d'exploitation sont stratégiques dans une logique concurrentielle, sur des marchés encore incertains. Pour prendre l'exemple de Bluely, aujourd'hui ils viennent de se lancer sur Lyon, leur modèle n'est pas encore mature, ils ne sont pas sûrs que ça va marcher aussi bien qu'à Paris. Et puis, ils ont eu l'expérience de l'espionnage par IBM. Aujourd'hui, s'ils mettent leur donnée en temps réel à disposition, quelqu'un peut tout enregistrer et retracer les parcours et l'usage des voitures. En récupérant ces enregistrements, les concurrents peuvent analyser leur modèle et proposer lors d'un prochain appel d'offre d'une collectivité une solution en s'appuyant sur ces retours d'expérience de Bolloré, sans en avoir eux mêmes fait l'expérience. Et il y a aussi le fait que ce soit un groupe coté en bourse. Si demain, quelqu'un, à partir des données, dit que le système ne marche pas du tout, et que le cours de la bourse baisse de 10%, c'est vraiment problématique pour eux. Donc, je comprends leur frilosité. [...] Je pense qu'ils attendent que Bluely soit vraiment installé dans le paysage lyonnais, et qu'il y ait un retour d'expérience de l'ordre d'un ou deux ans. Et que ce soit reconnu comme un succès, comme ça peut l'être à Paris. Aujourd'hui, en phase de lancement, c'est une phase de lancement ! C'est pas encore l'euphorie d'un service inscrit dans la durée. Je pense que la question se reposera à ce moment là. Velov au démarrage, il y a trois ans, quand on a parlé de données, ils étaient là, mais c'est les nôtres, on vous les donnera jamais, impossible, etc. Et aujourd'hui, c'est les premiers partenaires de SmartData.

On avait assez peu de crainte concernant l'espionnage par la concurrence à partir de ces données. D'une part, les données de disponibilité des stations étaient déjà diffusées sur notre site, et on savait qu'on avait un degré de sécurité assez faible, donc n'importe quel concurrent aurait pu les prendre. D'autre part, il y a d'autres moyens d'espionnage concurrentiel, une entreprise aurait pu

¹ Source : Entretien avec Guillaume, direction de la voirie, Grand Lyon (22 janvier 2014)

² Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

par exemple financer un thésard ou un cabinet d'étude pour analyser nos données d'exploitation sur les trajets que l'on fournit dans ce cas¹.

*

* *

L'épreuve de publicisation est une situation d'incertitude au cours de laquelle la donnée est questionnée : quel est son propriétaire ? Est-elle de bonne qualité ? Est-elle sensible ? A-t-elle un intérêt ? Est-on responsable en cas de mauvais usage ? Les réponses à ces questions ne sont pas catégoriques. Elles n'émergent qu'en relation avec des scénarios d'usage au cours desquels sont construits des « êtres intermédiaires ». Ces utilisateurs imaginés prennent des figures multiples (développeurs, utilisateurs malveillants, entreprises, militants, citoyens aux compétences techniques réduites, etc.), mais leur identité reste vague. Jamais définis précisément, ces premiers « êtres intermédiaires » n'incitent pas vraiment les producteurs à la diffusion de leurs données. L'activité de détachement passe par la remise en cause de ces cadres d'évaluation défavorables à l'ouverture des données afin d'entrouvrir la possibilité d'autres façons de l'aborder. De nouveaux êtres intermédiaires vont alors être associés aux données afin de parvenir à intéresser et à enrôler les producteurs à l'ouverture de leurs données.

C. Enrôler les producteurs : les dispositifs d'intéressement

Pour enrôler les producteurs, il est nécessaire de les intéresser, c'est-à-dire de susciter chez eux un intérêt à la diffusion de leurs données en stabilisant l'identité des données et des réutilisateurs². Pour parvenir à mobiliser ces acteurs, les chefs de projet ont effectué plus d'une cinquantaine de réunions de présentation visant à « vendre » leur projet auprès d'élus, de directeurs, de producteurs et d'utilisateurs de données³. A posteriori, l'un de ces chefs de projet considère que cette phase d'enrôlement des acteurs est la difficulté principale de l'ouverture des données.

La difficulté la plus importante est l'obtention des accords des fournisseurs de données à participer à une démarche alors qu'ils n'ont pas forcément en tête les enjeux et que ce n'est pas dans leur stratégie. C'est une vraie difficulté d'aller trouver des partenaires et de leur montrer l'intérêt. Elle a été menée par beaucoup, beaucoup de lobbying et de démarchage. Harald a fait

¹ Source : Entretien Thomas, direction des systèmes d'information, JC Decaux (27 août 2014)

² Michel Callon définit l'intéressement comme « l'ensemble des actions par lesquelles une entité s'efforce d'imposer et de stabiliser l'identité des autres acteurs qu'elle a défini par sa problématisation ». M. Callon, « Éléments pour une sociologie de la traduction. », art cit.

³ Plus d'une cinquantaine de réunions sont effectuées auprès de producteurs, directions, utilisateurs, etc.

énormément de travail auprès des partenaires mobilité pour obtenir des données. [...] Le volet obtention de données, il est quand même compliqué, et ce n'est pas derrière nous, c'est quelque chose de permanent¹.

En effet, les producteurs de données perçoivent rarement l'intérêt à mettre à disposition leurs données. « *Au début, ils ne comprenaient pas très bien pourquoi ils diffuseraient, à quoi ça servirait, donc, on a vu des mails passer en disant : « pourquoi je suis obligé de les diffuser ? » »*².

C'est un projet qui s'impose aux producteurs alors qu'ils ne l'ont pas demandé. Il peut y avoir aussi des réticences qui viennent de ça aussi, tout bêtement. [...] On leur dit il faut publier des données sur SmartData, je veux dire, c'est pas leur cœur de métier, donc ça passe après. S'ils n'y voient pas un intérêt et qu'ils n'ont pas de demandes politiques pour le faire, ça passe après le reste. [...] Ils gèrent leurs données métiers pour, à la direction de l'eau, c'est pour gérer le réseau d'eau. C'est pas pour faire autre chose, c'est pas pour s'amuser et pour diffuser de l'information. Ils ont besoin de ça, c'est un outil de gestion au quotidien pour les gugusses qui font du curage, pour l'entretien, pour tout un tas de bonnes raisons, mais pas du tout pour faire de la diffusion. C'est la énième roue du carrosse³.

Les données sont avant tout perçues comme des outils de travail quotidien des administrations. Pour faire émerger les conditions de possibilité de leur diffusion, il est alors nécessaire de traduire les données et d'associer dans un même mouvement les producteurs à de nouveaux utilisateurs. Plusieurs dispositifs d'intéressement⁴ permettent de mobiliser ces acteurs hétérogènes dans une démarche d'intérêt commun : des exemples d'utilisation, une plateforme facilitant leur besoin de diffusion, la valorisation du travail des producteurs, des licences leur permettant de conserver une maîtrise de leurs données, etc.

1) Réduire l'incertitude par la mise en avant d'exemples d'utilisation

L'*open data* présuppose une diffusion des données sans préjuger des usages qui peuvent en être fait. Cette dissociation des données et des usages provoque une incertitude chez les producteurs de données dont les interrogations à l'exposé du projet *open data* sont invariablement : « *Qui sont les utilisateurs ?* » et « *Quels usages vont être effectués de mes données ?* ». Les réponses initiales des porteurs du projet sont vagues, indiquant que les

¹ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

² Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

³ Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

⁴ Madeleine Akrich, Michel Callon et Bruno Latour, « A quoi tient le succès des innovations ? 1 : L'art de l'intéressement », *Gérer et comprendre. Annales des Mines.*, 1988, n° 11, p. 4-17.

données ouvertes répondent à des objectifs de développement économique, de création de nouveaux services et de transparence de l'administration. Toutefois, rapidement, ils présentent des exemples concrets afin de réduire cette incertitude et susciter l'adhésion des producteurs.

On a beaucoup présenté le projet, on a multiplié les présentations dans les différentes instances. Mais on ne l'a pas fait sur le mode « *On ouvre nos données et on verra ce qui se passe, ce qu'il en sort* ». Ça, ça a plutôt tendance à faire peur aux directions. Il faut leur présenter l'open data en montrant des exemples de ce qui peut être fait à partir de leurs données, en leur montrant des services qui peuvent être créés¹.

On se retrouve néanmoins ici face à un paradoxe : comment identifier des usages concrets des données alors que l'accès aux données ne nécessite aucune identification ? Le peu d'exemples connus d'utilisation est une difficulté pour intéresser les services comme l'indique le chef de projet :

Pour arriver à garder mobilisées les équipes qui travaillent sur ce sujet-là; je trouve qu'avoir des retours de réutilisations et de success story sur le territoire, ça serait quand même vachement intéressant. [...] [De ne pas en avoir] c'est une difficulté à la fois pour valoriser le projet en interne, mais aussi aux décideurs, aux politiques, aux partenaires, quand on va voir des partenaires et qu'on leur dit on aimerait mettre vos données dans la plateforme, ils nous disent souvent : « *ok, mais qui les utilisent ? à quoi ça sert ? quels sont les succès rencontrés ?* » C'est difficile de répondre à ça. Là dessus il y a un progrès à faire. Et même en interne, dans les équipes, à chaque fois que l'on a un retour d'utilisateur qui nous dit j'ai utilisé vos données pour faire ci ou ça, je vois, ça rebooste tout le monde².

Des exemples d'utilisation provenant soit d'autres collectivités, soit de retours indirects sont cités par les chefs de projet. Un exemple récurrent est celui d'un transporteur de la région lyonnaise qui utiliserait la donnée des chantiers perturbants pour optimiser ses livraisons. Fiction ou réalité ? Lorsque je les interroge pour en savoir plus sur cette entreprise, personne n'est en mesure de m'en indiquer davantage. L'histoire est pourtant répétée inlassablement lors des réunions de présentation du projet. Ce « transporteur » devient un « porte-parole » de l'ensemble des utilisateurs de données ouvertes, mobilisé pour légitimer l'ouverture des données et intéresser les producteurs.

¹ Journal de terrain (20 janvier 2013)

² Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

Encadré 8 : Les hackathons, des dispositifs de mobilisation des utilisateurs ou d'intéressement des producteurs ?

Les politiques d'*open data* sont étroitement associées aux *hackathons*, des événements réunissant des développeurs afin de créer des applications. Après avoir été mis en œuvre dans des villes américaines¹, ces dispositifs ont été transposés par des collectivités françaises sous des formes diverses (concours, appel à projet, etc.). Ils visent à mobiliser des utilisateurs et à encourager le développement économique et la création de services à partir des données ouvertes². Cependant, la pérennité des services créés et les retombées économiques sont souvent décevantes : une large majorité de ces applications ne se sont pas traduites dans des succès commerciaux et industriels³. Pourtant, les institutions publiques continuent à organiser ces événements coûteux aux résultats incertains. Ce paradoxe s'explique par le fait que ces dispositifs ludiques et compétitifs ne visent pas tant à mobiliser des utilisateurs qu'à intéresser les producteurs. Ils participent à la construction d'un récit, qui légitime les politiques d'*open data*, en associant les données au développement économique. Les hackathons démontrent « l'existence d'une demande de données⁴ », en fournissant de nombreux exemples d'utilisation mis ensuite en avant pour inciter les producteurs à diffuser davantage de données.

2) Faciliter la diffusion des données en ayant recours à la plateforme open data

La plateforme de diffusion de données est un dispositif d'intéressement pour les acteurs par le fait qu'elle facilite la mise à disposition de leurs données et en assure leur qualité. Cette nouvelle infrastructure technique devient un « point de passage alternatif » à des situations

¹ Le premier concours a été mis en place par la ville de Washington en 2007. La ville de New York organise également annuellement les NYC Big Apps depuis 2009. Les retombées économiques annoncées de ces premiers événements ont justifié leur duplication dans d'autres villes. Alors que le coût de l'opération était de 50 000 dollars, la valeur des 47 applications créées lors du concours était estimée à 2,3 millions de dollars.

² Certains voient ainsi dans ces concours de nouvelles formes de mobilisation et d'innovation politique. Laurence Allard et Olivier Blondeau, *Nouvelles formes de mobilisation et d'innovation politique : le concours*, <http://www.internetactu.net/2010/06/10/nouvelles-formes-de-mobilisation-et-dinnovation-politique-le-concours/>, 6 octobre 2010, (consulté le 8 avril 2015).

³ Par exemple, Rennes Métropole a lancé un concours en octobre 2010 qui a donné naissance à 43 applications. La quasi-totalité de celles-ci n'existent plus aujourd'hui ou ne sont plus mises à jour. Ce cas n'est pas isolé, les applications développées dans le cadre des NYC BigApps survivent rarement plus que quelques mois après ces concours.

⁴ Samuel Goëta, « Des données à la recherche de publics: les assemblages temporaires des concours de réutilisation de données ouvertes », *Participations*, à paraître.

jugées peu satisfaisantes pour les acteurs. La plateforme est un « facilitateur obligé du détachement¹ ».

En premier lieu, le choix de l'*open data* est pour certains acteurs une réponse à des problématiques techniques. Ne diffusant pas leurs données, ils étaient l'objet de tentatives de récupération de leurs données sur leur site internet par des robots informatiques. Ces opérations leur posaient deux problèmes : elles mettaient en péril leur infrastructure technique, et elles ne garantissaient pas la qualité des données récupérées. Pour éviter l'utilisation « sauvage » de leurs données et en conserver la maîtrise, l'entreprise JC Decaux a par exemple fait le choix de mettre à disposition ses données de VLS sur une plateforme *open data*.

Le sujet de l'accès aux données des vélos en libre service s'est posé assez rapidement dans la mesure où 80% du trafic de nos sites internet étaient des robots qui venaient extraire les données pour les utiliser dans des applications ou des sites tiers. Ca nous posait des problèmes puisque ça nuisait à la bonne exploitation de nos sites. Certains de ses robots pouvaient être assimilés à des attaques sous forme de déni de service, qui saturent nos serveurs par les requêtes trop fréquentes. [...] L'entreprise a alors pris la décision de mettre ses données en open data, dans la mesure où l'on estimait qu'il s'agissait du sens de l'histoire et qu'il valait mieux accompagner la marée plutôt qu'essayer de lui résister².

La plateforme offrant un accès libre aux données est également perçue comme un moyen de réduire les coûts de diffusion des données. Les acteurs n'ont plus à répondre aux demandes régulières particulièrement chronophages.

Nous, la diffusion, j'ai presque envie de dire que ça s'est fait naturellement. On mettait beaucoup de données à disposition de bureaux d'études ou de partenaires. Ca nous prenait pas mal de temps. Diriger ces demandes vers un portail c'est finalement plus confort pour nous. Ouais, [L'*open data*] c'est du boulot en moins aussi, parce que du coup, t'as pas d'exports à faire pour une boîte qui pourrait te demander tes données. [...] Donc tout le monde y a gagné : les prestataires, parce qu'il y a quelque chose de structuré sur SmartData, et nous parce qu'on gagne du temps³.

La plateforme répond ainsi aux besoins des services en facilitant l'échange de données entre acteurs. Elle est également nécessaire pour des projets innovants mis en place par l'infrastructure telle que la Centrale de mobilité ou la modernisation de la diffusion des données aux communes (cf. chapitre 2). Les producteurs perçoivent alors directement l'intérêt de la plateforme et ne rechignent pas à la diffusion de leurs données.

¹ F. Goulet et D. Vinck, « L'innovation par retrait. », art cit.

² Source : Entretien Thomas, direction des systèmes d'information, JC Decaux (27 août 2014)

³ Source : Entretien Martin, administrateur de données, Grand Lyon (6 décembre 2013)

Enfin, la plateforme permet également de diffuser le plus largement possible ses données afin qu'elles deviennent un référentiel partagé par l'ensemble des acteurs.

Et alors, là, pour le coup, j'ai aucun doute sur le fait que, le fond de plan, c'est de la donnée qui doit être diffusée le plus largement possible. [...] Ca permet de faire de la carto facilement, ça permet de faire en sorte que tous les utilisateurs potentiels du SIG partagent un même référentiel. Quelque part, c'est tout bénéf pour tout le monde, car quand on échange les choses se superposent au mieux si tout le monde s'appuie sur le même référentiel, on est tous d'aplomb. Il y a un vrai enjeu par rapport à ça, à diffuser au plus large nos informations, de manière à ce que le SYTRAL ou les communes, tous nos partenaires s'appuient sur le même référentiel que nous¹.

Cette diffusion, la plus large possible, des données de la communauté urbaine permet également d'éviter un autre risque perçu : que la donnée soit produite par des acteurs externes et qu'ils perdent la maîtrise de représentation de l'espace urbain.

Si tu ne diffuses pas, tu peux avoir la même personne, qui décide de créer elle-même sa base de données et dans ce cas là, c'est encore pire. T'imagines quelqu'un qui dit « *ils ne mettent pas les silos, je vais le faire moi-même. Je vais aller faire une campagne, je vais aller relever tous les silos.* » Mais après, la mise à jour, elle est faite, tu sais pas quand, c'est pas possible, il vaut mieux que tu la diffuses et tu vois ce qui se passe².

3) Valoriser le travail des producteurs

L'*open data* s'inscrit également dans une stratégie de communication et joue sur l'image de l'institution productrice. La réutilisation des données permet de valoriser le travail des producteurs, qui peuvent s'en servir pour justifier le financement de leur activité auprès de la direction de la communauté urbaine.

Le côté positif, c'est « *wahouh, le Grand Lyon, c'est chouette ce qu'ils font* ». Et là, les gens ils vont faire de la com, de la pub, ils vont voir les élus, ils vont dire « *c'est génial ce que vous faites, encore, encore* ». On peut le voir aussi comme ça. Je préfère le voir comme ça d'ailleurs. Si on a de bons échos, de pleins de services, de prestataires, ils mettront encore plus d'énergie et de moyens pour qu'on produise ces données³.

Après si on peut justifier de l'intérêt que le public porte à nos données, ça consolidera auprès de la direction en interne, et même auprès du Grand Lyon de manière plus générale, le côté innovant, dynamique et moderne de la direction de la voirie. [...] donc ça participe à une valorisation et à une consolidation des moyens affectés à nos missions. Parce que derrière, dans le rapport d'activité, Voirie 2013, il y a la partie voirie classique, le nombre de potelets, de passages piétons, et il y a quand même des paragraphes, notamment portés par VMU, d'innovations, de nouveautés,

¹ Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

² Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

³ Source : Entretien Romain, administrateur de données, Grand Lyon (21 novembre 2013)

on parle de SmartData, enfin c'est aussi une fierté de la direction, de montrer que l'on est présent de ce point de vue là¹.

En l'absence de connaissance des utilisateurs, cette valorisation se matérialise par les compteurs sur le site indiquant le nombre de consultation de leurs données, que les services mettent en avant dans leurs rapports annuels d'activité.

Les compteurs, c'est intéressant pour nous de savoir, parce qu'on se défonce pas mal pour faire ces bases de données, essayer de les enrichir, on met beaucoup d'énergie, et c'est vrai qu'on ne se rend pas trop compte de l'impact de l'utilisation ou pas des données. La reconnaissance, on l'a quand on voit les compteurs qui montent².

La responsable de la plateforme présente mensuellement aux producteurs un Top 3 des données les plus consultées. En mettant en concurrence les services, ces formes de benchmarking visent à accentuer la publication de données sur la plateforme *open data*.

Le caractère innovant associé à l'*open data* permet également de renforcer les stratégies de communication des acteurs mettant à disposition leurs données. Il vaut mieux en être plutôt que de se voir accoler une image « passiste » et de marcher à contre-courant.

Aujourd'hui, il y a tout le travail qui a été fait sur la ville intelligente, qui intègre SmartData, qui nous aide à avancer là dessus. En gros, on a tellement rendu visible cette démarche ville intelligente qui contient tous ces petits projets, du coup les partenaires adhèrent plus facilement à ces projets parce que ça s'inscrit dans une démarche plus globale et plus visible. Il vaut mieux en être que bloquer le projet ville intelligente du président du Grand Lyon³.

A contrario, l'image négative est un moyen de pression pour les acteurs réticents. Ainsi, à propos d'un acteur n'ayant pas souhaité mettre ses données à disposition, le responsable du projet Optimod me déclare : « *Je leur ai juste fait une remarque en leur disant qu'ils étaient les seuls sur la place de Lyon à ne pas avoir signé. Et qu'on le ferait savoir⁴ !* ».

4) Les licences comme dispositif d'intéressement

Enfin, les licences encadrant l'utilisation des données participent à l'intéressement des producteurs réticents à les diffuser en accès libre. En proposant une ouverture graduée selon trois modalités, elles permettent à ces acteurs de garder un contrôle sur leurs données. Lors

¹ Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

² Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

³ Source : Entretien avec Guillaume, direction de la voirie, Grand Lyon (22 janvier 2014)

⁴ Source : Entretien avec Harald, chef de projet, Grand Lyon (8 janvier 2013)

des échanges avec les opérateurs de transport, le responsable du projet Optimod met ainsi constamment en avant les trois modalités envisageables d'ouverture de leurs données. Selon lui, ces licences ont été un facteur majeur de succès d'ouverture des données des « partenaires » de la communauté urbaine, en particulier dans le secteur de la mobilité.

Je pense que l'on a rassuré en fait. On a rassuré parce qu'avec notre stratégie de mise à disposition, on disait, on ne met pas à disposition de n'importe qui, si vous voulez vous pouvez savoir à qui vous mettez à disposition, vous pouvez voir les services qu'il fait. Si vous n'êtes pas d'accord, vous nous le dites. Et ça, ça a quand même pas mal rassuré les gens. Je pense qu'on y aurait été open data bille en tête, il y a un paquet de gens qui auraient dit non. Et je crois qu'on aurait attaqué comme ça, on serait revenu derrière, on aurait ramé pour les convaincre de nos intentions. Donc, en fait, le fait d'adresser à la fois les gens en situation de monopole plus les réutilisations on va dire fantaisiste, ça a plutôt rassuré les gens, et c'est ce qui fait que ça a marché, je pense en fait. Moi je l'analyse comme ça a posteriori¹.

En réduisant les incertitudes sur les réutilisations, les licences sont un dispositif central d'intéressement et d'enrôlement des producteurs. En écartant les « êtres intermédiaires » indésirables, elles offrent aux producteurs une garantie de bénéficier des avantages de l'ouverture des données, tout en réduisant les risques d'usages malveillants.

*

* *

Ces dispositifs d'intéressement parviennent à réduire les craintes du détachement des données, sans toutefois amener une certitude sur le bénéfice de la diffusion des données. L'incertitude reste présente, mais elle est minorée en étant insérée dans un ensemble d'attachements. De nouveaux « êtres intermédiaires » sont définis au travers de ces dispositifs : des entreprises développant des services à partir des données ouvertes, des utilisateurs venant se servir librement et de manière sécurisée sans importuner les producteurs, etc. L'identité des données ouvertes est également clarifiée : elles sont stabilisées comme des opportunités à la fois de création de nouveaux services et de valorisation de l'image des institutions les mettant à disposition.

D. Qui décide de l'ouverture des données ?

Toutes ces questions sur la publicisation des données reçoivent des réponses hétérogènes et contradictoires. Une multiplicité de représentations d'utilisateurs des données est apparue sans qu'aucune ne s'impose réellement. Comment les acteurs parviennent-ils à se

¹ Source : Entretien avec Harald, chef de projet, Grand Lyon (8 janvier 2013)

coordonner face à cette prolifération ? Comment décider de la diffusibilité d'une donnée face à ces utilisateurs aux identités contradictoires ? Quels sont les êtres intermédiaires les plus consistants ? Et qui décide alors de leur attribuer une identité stabilisée ? En bref : qui décide de l'ouverture des données ?

Excepté pour certaines données concernées par la directive INSPIRE, les acteurs n'ont aucune obligation à publier leurs données. Les procédures de décision de l'ouverture des données sont alors très variables selon les directions. Il n'existe pas de schéma décisionnel commun au sein de la communauté urbaine, chaque producteur reste responsable de ses données et décide de leur publication. Certains prennent unilatéralement la décision de mettre à disposition les données.

Les décisions de publication ou pas, je les ai portées directement. On n'est pas remonté en comité de direction pour... parce que c'est des choses très techniques et très monde cartographique pour décider si on montre tel ou tel attribut, telle couche, telle donnée. C'est moi qui ai fait les choix. Si on avait enclenché cette démarche de demander l'analyse précise, ça aurait pris un certain temps, et aujourd'hui, on aurait pas proposé autant de couches à disposition¹.

Néanmoins, le choix est le plus souvent effectué après un passage en comité de direction, dont la décision est très variable selon les acteurs. Alors que la délégation au développement urbain a opté pour une ouverture large, le processus a été plus incertain au sein de la direction de la propreté avec « *pas mal d'allers-retours, de contre-ordres, de contre-indications* ».

Comment on valide le passage de nos données vers SmartData ? J'ai presque envie de dire, que quand nous on présente ces données là à notre directeur, il nous dit c'est bon allez-y. Il en a parlé en comité de délégation, mais mettez tout. [...] On a plutôt eu comme directive de mettre pratiquement toutes les données sur SmartData. J'ai été le voir avec la liste de nos données, ce qu'on en faisait aujourd'hui, qui l'utilisait, Présence SmartData Oui/Non. Et finalement, on a eu que du oui².

Ces procédures de validation décentralisées au sein de chacune direction, sans aucune directive commune à l'ensemble de l'institution, soulèvent des incertitudes. En l'absence de contraintes et de règles codifiées, un acteur peut décider unilatéralement de publier ou de ne pas publier les données qui sont sous sa responsabilité. Ainsi, une donnée publiée a été retirée de la plateforme en toute discrétion par le directeur d'un service de la communauté urbaine au grand dam des chefs de projet : « *Il faut un minimum de cohérence, pas comme avec les silos*

¹ Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

² Source : Entretien Martin, administrateur de données, Grand Lyon (6 décembre 2013)

à verre qu'ils ont retiré de la plateforme du jour au lendemain. Il faut que la mise à disposition soit pérenne afin que les réutilisateurs aient une certaine sécurité quant à leur utilisation des données¹. » En outre, en l'absence de directives précises, certains acteurs ne savent comment se positionner. Ils souhaiteraient des orientations de la part des élus.

C'est là où aujourd'hui, je trouve, on manque cruellement de gouvernance autour de SmartData, c'est qu'on a pas de directives politiques pour dire « *Oui, on a plutôt envie de faire payer des redevances pour récupérer un peu de budget* », ou alors au contraire, « *non politiquement, on estime que c'est de l'information financée par des budgets déjà publics et qu'il n'y a aucune raison de se faire de la gratte autour de ce truc là* ». Je sais pas ils pourraient décider l'un ou l'autre, et nous on fera. Mais, du coup, comme on a pas ce niveau de gouvernance, ben, on le fait à notre sauce. [...] Mais aujourd'hui, on fait valider ça à qui ? Non seulement on a pas de directives qui viennent du dessus, mais on a même pas d'endroit pour nous faire donner une réponse par rapport à ça. A tout le mieux, ce qu'on peut avoir, c'est dans un comité de pilotage, de dire, « *ben on fait ça, ça vous va ?* » Et ça je trouve que ça manque. Est-ce qu'il va se monter quelque chose ou est-ce que c'est perdu d'avance ? Est-ce les techniciens qui vont décider ? C'est ce qui se passe pour l'instant².

La décision est ainsi aux mains de l'administration, ce qui n'incite pas à des processus de publication rapides des données.

Et à mon sens, tant que l'on n'aura pas le petit déclic politique pour pousser SmartData, on ira guère plus vite qu'à ce rythme de sénateur. Qui avance mais pas très vite. Si demain, en 2014, on a un vice-président, comme le gars de Saône et Loire, qui dit « *je veux qu'on soit exemplaire sur les données de SmartData* », ça ira plus vite. Parce que les directeurs vont pousser, le MOI, il va pousser, l'administrateur de données, il mettra ça en haut de la pile³.

Cette question soulève également des enjeux en cas de conflit entre plusieurs directions quant à la publication des données.

Après, on avait une question de gouvernance, c'est la première question qui est arrivée en disant « *Oui, mais imaginons que l'on ait la même donnée à l'eau et à la propreté, et qu'il y a une direction qui veut la diffuser mais pas l'autre. Qui est ce qui décide ?* » C'était la première question qui était ressortie. [...] Mais ce qui m'avait interpellé, c'est que la première question était une question de gouvernance : qui est-ce qui décide de la donnée, au bon moment, par rapport à l'ensemble du Grand Lyon. Question qui est toujours là⁴.

Pour limiter ces incertitudes, des schémas de publication, dotés d'instances de gouvernance de l'ouverture des données, sont proposés par l'équipe projet, mais ils ne seront jamais mis en place. Les questions soulevées dans le paragraphe introductif n'ont ainsi aucune

¹ Source : Journal de terrain, Clément, chef de projet

² Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

³ Source : Entretien Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

⁴ Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

réponse précise. Dès lors, impossible de préciser clairement les attributs du « public » des données ouvertes : celui-ci est relatif à chacune des données mises à disposition.

Encadré 9 : La mise à disposition des données du SYTRAL

Le plus dur ça a été les données TCL, là vraiment, on a du... TCL, ça a été une vraie galère, 5, 6 notes au Président, des rappels avec le cabinet qui ne voulait pas y aller trop fort non plus. Alors qu'en fait tous les services étaient d'accord en disant : « *il faut juste que Gérard et Bernard se mettent dans un bureau et qu'ils se mettent d'accord. C'est tout* »¹.

Les particularités du travail d'enrôlement du SYTRAL requièrent une présentation spécifique du processus de mise à disposition de ses données. Très demandées par les réutilisateurs, les données de transport en commun sont souvent parmi les premières à être publiées par les collectivités². Pourtant, les données du SYTRAL, indispensables au projet de Centrale de mobilité de la communauté urbaine de Lyon n'ont été mises à disposition qu'en février 2015³. Pourquoi plus de trois ans de négociations ont été nécessaires ? Ce délai est d'autant plus surprenant que dès le printemps 2013 les équipes administratives de la communauté urbaine et de l'autorité organisatrice des transports urbains travaillent conjointement pour permettre cette diffusion.

Contrairement à d'autres AOT françaises (Rennes, Bordeaux, etc.), ce n'est pas l'exploitant du réseau – ici Kéolis – qui est propriétaire des données mais bien le SYTRAL. Ce dernier a donc le pouvoir de mettre à disposition les données de transports en commun. Pourquoi ne voulait-il donc pas ouvrir ses données ? Les justifications avancées sont plurielles, et rejoignent en grande partie les craintes des producteurs présentées précédemment : valorisation des données par des acteurs commerciaux, souci de qualité des données et un intérêt limité pour l'*open data*.

Il y avait trois arguments. Un : « *Je ne veux pas que des opérateurs privés se fassent de l'argent sur le dos de mes données* ». Donc, en gros, ils ne partagent pas du tout l'objectif open data qui est complètement l'inverse : profitons des données pour faire en sorte qu'il y ait de l'activité économique qui se développe dessus. Je le traduis comme ça. Deuxième point, c'était : « *ça va nous imposer un travail de qualification de la donnée, ça va mobiliser des moyens pour être sûr que la donnée soit de qualité* ». Là-dessus, la donnée, ils la jugent déjà de qualité puisqu'ils s'appuient dessus pour développer un service en propre. Pourquoi est-ce qu'elle ne serait pas de qualité suffisante pour la confier à un opérateur privé alors qu'ils l'utilisent déjà pour leur propre outil ? Et au pire, c'est la seule donnée qui est disponible. Personne n'aura une donnée plus propre que celle-là. Si eux, ils n'arrivent

¹ Source : Entretien avec Harald, chef de projet, Grand Lyon (19 janvier 2015)

² Les politiques d'open data de plusieurs villes américaines (New York, Philadelphie) et françaises (Rennes) trouvent leur origine dans la mise à disposition des données des transports en commun.

³ Les données mises à disposition sont des données géographiques (localisation des points d'arrêt et stations, tracé des lignes, localisation des parcs relais, des agences commerciales, etc.), les horaires théoriques aux arrêts de l'ensemble des lignes et des données temps réel (temps d'attente des prochains passages aux arrêts, perturbations importantes inopinées sur le réseau, disponibilité en temps réel des équipements relatifs à l'accessibilité).

pas à le faire, personne d'autre n'y arrivera non plus. Donc autant, la mettre à disposition, c'est celle qui existe et la plus qualitative, même si elle a des défauts, mais comme n'importe quelle donnée, on n'arrivera jamais à un truc parfait. Et enfin, troisième point, c'était : « *de toute façon, ça ne servira à rien, le service que vous allez déployer en terme de multimodalité, il n'aura pas d'intérêt, ça ne va pas faire changer le comportement des gens* »¹.

En outre, alors que le SYTRAL vient de lancer sa propre application mobile², il craint de subir une perte de visibilité si ses données sont intégrées dans des services proposés par des acteurs tiers. Enfin, et surtout, au travers de la mise à disposition des données se joue un enjeu de maîtrise de l'information fournie aux usagers et de positionnement de l'institution dans la gouvernance urbaine. « *Or, l'ouverture des données, [le président du SYTRAL] vit ça comme une perte de pouvoir, d'où son refus catégorique. Il ne veut pas perdre le contrôle, en fait il veut avoir le monopole de l'info TCL*³. »

Néanmoins, parallèlement à ces discours, les services administratifs du SYTRAL donnent un accord de principe à la diffusion de leurs données sur la plateforme de la communauté urbaine. Dès le printemps 2013, ils travaillent alors conjointement avec les équipes du Grand Lyon sur les modalités juridiques et techniques de cette mise à disposition. Les licences sont rédigées avec les services juridiques du SYTRAL. Il y a ainsi « *un décalage entre le discours politique et le discours technique. Les techniciens sont tous d'accord pour y aller, mais ils n'ont pas forcément l'accord explicite au niveau politique pour une diffusion à large échelle*⁴. » Ainsi, ce n'est plus le SYTRAL qui ne veut pas mettre à disposition ses données, mais son président, qui s'oppose à l'*open data*.

Comment l'ouverture des données du SYTRAL a alors été obtenue par le Grand Lyon ? Contrairement aux cas précédents, l'enrôlement du SYTRAL ne s'est pas tant effectué par des dispositifs d'intéressement que par des dispositifs de contrainte. La diffusion des données de transport en commun a été imposée au président du SYTRAL par le président de la communauté urbaine. Il s'agit du seul cas où les élus sont intervenus directement dans le processus d'ouverture des données. Cette intervention politique fait suite à une impasse des services administratifs. En avril 2013, lors d'un colloque organisé conjointement par le GART

¹ Source : Entretien avec Guillaume, direction de la voirie, Grand Lyon (22 janvier 2014)

² « *Les données temps réel des transports, on n'est pas sûr de les ouvrir. On vient de créer nos applications smartphones, on ne voit pas trop à quoi cela servirait de les ouvrir à la concurrence* ». (Geoffrey, Journal de terrain, 15 février 2013). L'application mobile TCL a été lancée en janvier 2013.

³ Source : Entretien avec Harald, chef de projet, Grand Lyon (19 janvier 2015)

⁴ Source : Journal de terrain (20 mars 2013)

et le SYTRAL sur « La gestion de l'information liée à la mobilité, d'un enjeu stratégique à un agir politique », le président du SYTRAL, Bernard Rivalta affirme publiquement son opposition à l'*open data*. Cette posture très restrictive est perçue par les agents du Grand Lyon comme une menace pour l'image de l'agglomération à quelques jours du lancement officiel de la plateforme SmartData. « *A mon sens la posture de non ouverture par principe ressemble à une ligne Maginot, et ne tiendra pas dans le temps. [...] En plus, sa position fermée, ça donne vraiment une mauvaise image pour l'agglomération, qui rejaillit sur le Grand Lyon. Il faut qu'on intervienne au niveau du cabinet [du président] du Grand Lyon pour qu'ils lui fassent changer de position* »¹. Une note est alors adressée au cabinet du président de la communauté urbaine afin que celui-ci intervienne pour faire converger les points de vue entre les deux institutions. Un an plus tard, alors que la campagne électorale fait rage, les positions du président du SYTRAL n'ont pas évolué. Or, Gérard Collomb souhaite s'appuyer sur un des projets phares de la démarche de ville intelligente pour défendre son bilan. « *Collomb m'a appelé pendant la campagne pour l'appli Optimod, etc., et je lui ai dit : « tant que je vous ai au téléphone, ce qui ne m'arrive pas souvent, je vous le dis tout de suite, l'appli Optimod, elle ne se fera pas si on a pas les données TCL. Pas de données TCL, pas d'Optimod. » Alors là, ça a fait boum. (rires)* »². La notoriété du projet Optimod et les enjeux politiques qui en découlent ont alors contribué à la mise à disposition des données de transport en commun. « *Il vaut mieux en être que bloquer le projet du président, ville intelligente. Parce que ça pousse au niveau politique. Toute la publicité, la promotion qui s'est faite autour d'Optimod bénéficie à ces services-là. Si on n'avait pas eu ce retour média sur Optimod, jamais on n'aurait fait ça. On n'avancerait pas. Le SYTRAL nous laisserait patouiller dans notre coin à faire des expérimentations. La com, ça permet de faire avancer, ça donne une nouvelle importance* »³. La diffusion des données de transports en commun va en effet être actée quelques semaines plus tard lors d'une réunion entre les présidents du Grand Lyon et du SYTRAL.

Pour les données SYTRAL, c'est simple, le sujet a été remis une nouvelle fois sur la table en DGCab après que l'on ait inondé de mails et de notes le cabinet. Peu de temps après, il y a une réunion Grand Lyon / Sytral, avec pas mal de sujets importants, et des sujets à la con, comme la mise à disposition des données. Collomb prend la parole au milieu de cet ordre du jour "Pour l'ouverture des données, c'est

¹ Source : Journal de terrain, Harald (5 avril 2013)

² Source : Entretien avec Harald, chef de projet, Grand Lyon (19 janvier 2015)

³ Source : Entretien avec Guillaume, direction de la voirie, Grand Lyon (22 janvier 2014)

bon ?" et le président du Sytral lui répond "Je ne suis pas d'accord, mais c'est bon". Au milieu de tous les sujets, il a du laisser filer là dessus¹.

Cette lutte pour obtenir l'accord de diffusion des données de transport en commun révèle les relations particulières entre le Grand Lyon et le SYTRAL. Si les élus du Grand Lyon sont majoritaires au sein du comité syndical et que le président du SYTRAL a toujours occupé parallèlement un mandat de vice-président de la communauté urbaine, les deux institutions sont autonomes². Seconde autorité organisatrice des transports urbains de France, le SYTRAL défend farouchement son indépendance et son leadership sur la politique de transport en commun. Or, Les négociations pour l'ouverture des données du SYTRAL se sont déroulées parallèlement au processus de création de la métropole de Lyon. Ce contexte de changement institutionnel a pu soulever des craintes d'une perte d'autonomie de l'AOTU vis-à-vis du Grand Lyon³. En février 2015, les données sont ouvertes quand Bernard Rivalta n'est plus président du SYTRAL et que Gérard Collomb en occupe provisoirement la présidence.

*

* *

L'épreuve de publicisation souligne le lien étroit entre la donnée et son usage. Contrairement aux discours des militants de l'*open data* qui incitent les producteurs à diffuser leurs données sans se préoccuper des usages qui en découleront, en pratique, le processus d'ouverture des données se caractérise par la construction d'utilisateurs imaginés. Ces « êtres intermédiaires » prennent des formes multiples selon les données et les enjeux auxquels elles sont associées. Cette phase de publicisation nous révèle, dans la continuité des travaux des pragmatistes américains⁴, qu'il n'existe pas un public homogène. Il y a plusieurs publics

¹ Source : Journal de terrain (11 juillet 2014)

² Cela se manifeste lors des réunions communes entre le Grand Lyon et le SYTRAL auxquelles j'ai pu assister. Les membres des deux institutions ne se mélangent pas, chacun reste entre pairs de part et d'autre des tables, et un certain formalisme est respecté lors des échanges. Comme je le note dans mon journal de terrain : « *on remarque la distinction entre ces deux entités au cours de la réunion, au travers d'échanges, qui mettent en avant la différence entre les deux institutions et la faiblesse des échanges entre elles, mais aussi plus prosaïquement dans l'agencement des individus au cours de la réunion, chaque organisation étant séparée bien distinctement autour de la table.* » (Source : Journal de terrain, mars 2015)

³ En témoigne la remarque d'un agent de la direction de la voirie du Grand Lyon lors d'un comité de pilotage du projet SmartData : « *Dans le contexte de création de la Métropole, il faut faire attention aux susceptibilités. Aujourd'hui, le Sytral sait qu'il risque de se faire manger d'ici deux ans par le Grand Lyon, donc, autant dire que les rapports avec la direction de la voirie sont difficiles et tendus !* ». (Source : Journal de terrain, 21 janvier 2013).

⁴ W. Lippmann, *The phantom public*, op. cit. ; J. Dewey, *Le public et ses problèmes*, op. cit.

« fantômes » pour reprendre l'expression de Lippmann qui sont rassemblés autour d'un problème (*issue*) spécifique matérialisé ici par une donnée. En aucun cas, il n'est possible d'uniformiser ce public. Dès lors, dans l'ouverture des données publiques, la notion du « public » est tout autant à interroger que celle de la « donnée ». Plutôt que l'ouverture des données publiques, il faut ainsi questionner l'ouverture des données aux publics.

En corollaire de la mise en visibilité des données sur une plateforme publique, l'épreuve de publicisation révèle également les dynamiques inverses de constitution d'opacité autour des données non publiées. L'ensemble des questionnements auxquelles sont soumises les données, leur association à de nouveaux usages et utilisateurs, les incertitudes des producteurs, les scènes de négociation et d'intéressement constituent un processus de mise en visibilité tout autant qu'ils rendent invisibles les données non publicisées à l'issue de cette épreuve. Solidement attachées à leurs producteurs, ces dernières n'acquièrent pas les caractéristiques de données diffusibles.

III. Extraire

La dernière épreuve de diffusibilité de la donnée est son extraction. Elle correspond à l'ensemble des opérations nécessaires pour extraire la donnée de son environnement initial et la mettre à disposition sur une infrastructure de diffusion. Les données ne sont en effet jamais indépendantes d'une infrastructure technique au sein de laquelle elles sont produites et utilisées. Ces systèmes d'information n'ont pas toujours été conçus pour permettre l'extraction et la mise à disposition des données qu'ils contiennent. Une transformation de l'infrastructure est souvent réalisée pour permettre la mise à jour automatique des données sur le portail, un changement de format de données ou la diffusion de données volumineuses et en temps réel, tout en assurant la sécurité du système d'information de l'institution. Ces opérations mettent à l'épreuve les liens qui associent la donnée à un système technique et participent au travail de détachement de la donnée de son environnement initial et de son attachement à un système externe. Cette épreuve doit permettre de stabiliser l'identité des données comme « données ouvertes ».

A. Extraire la donnée

La première épreuve consiste à extraire la donnée du système d'information du producteur afin de la mettre à disposition sur un portail de diffusion. Cette opération vise à détacher la donnée de son infrastructure initiale pour l'associer à une autre infrastructure. L'extraction est une médiation qui permet tout autant le détachement que l'attachement de la donnée. Elle peut paraître purement technique et formelle, pourtant, elle nécessite la mise en place d'une organisation sociale et détermine la nature du lien entre le producteur et l'utilisateur de la donnée. Un public est ainsi préfiguré au travers des deux modalités d'extraction de données envisageables : manuelles ou automatiques.

1) L'extraction manuelle

L'extraction manuelle repose sur le travail d'opérateurs : une personne extrait la donnée du système d'information puis la transfère à intervalles réguliers sur la plateforme de diffusion de données. De nombreux portails *open data* ont été initialement conçus de cette façon qui présente l'avantage de la simplicité. Seulement, l'extraction manuelle présente l'inconvénient d'être chronophage, notamment lors des phases de mises à jour des données. Comme le souligne la responsable du portail *open data* de la région PACA, il est nécessaire de contacter à intervalles réguliers chacun des producteurs pour obtenir des données actualisées.

Il n'y a pas de moissonnage automatique. Les partenaires m'envoient leurs données, je remplis la fiche de métadonnée et je les publie. La mise à jour des données se fait manuellement. Il s'agit principalement de données dont la fréquence de mise à jour est annuelle. Là je me concentre d'abord sur la mise à jour de nos données en interne et ensuite, j'irais réclamer les données chez chacun de nos partenaires. [...] C'est chronophage, c'est l'équivalent d'un temps plein. Le principal poste de dépense de l'open data, c'est celui-là, pas le portail dont le coût est relatif. Et, ça nécessite beaucoup de relationnel avec les différentes personnes à contacter, à recontacter...¹

Excepté pour une donnée rarement actualisée, le processus de publication de la communauté urbaine de Lyon ne repose pas sur l'extraction manuelle qui complique les opérations de mise à jour. L'extraction manuelle rompt le lien entre la base de données du producteur et la plateforme de diffusion : si le producteur met à jour ses données, l'actualisation n'est pas automatique sur la plateforme. Deux données similaires coexistent alors de manière autonome, la donnée utilisée par les services producteurs et la donnée ouverte, au risque que cette dernière, si elle n'est pas actualisée, ne soit pas jugée fiable par

¹ Source : Entretien téléphonique avec Gabrielle, responsable open data de la région PACA (15 mai 2013)

les réutilisateurs. Cela révèle la tension entre attachement et détachement qui traverse l'ensemble du processus d'extraction. Pour être catégorisée de manière pérenne comme « donnée ouverte », la donnée doit être suffisamment détachée de son infrastructure initiale pour pouvoir être attachée à une infrastructure tierce, sans être trop détachée au risque que la nouvelle association ne soit pas pérenne. La mise en œuvre de procédures d'extraction automatique des données résout cette tension.

2) L'extraction automatique

L'extraction automatique repose sur une chaîne de traitement informatique appelée ETL (*Extract, Transform & Load*). Un ETL extrait les données, les transforme et les charge dans une base de données. Il s'agit d'un connecteur qui « moissonne » la donnée depuis la base de production jusqu'à la base de diffusion à une fréquence donnée (automatiquement ou à la demande, chaque année/mois/semaine/jour selon le type de données). Les ETL permettent d'industrialiser le développement et l'exécution de chaînes de traitement de données. La mise en place de ces chaînes d'extraction nécessite un travail conséquent sur le système d'information afin d'assurer la fiabilité du « moissonnage » et la sécurité de l'infrastructure de production¹ comme le décrit un responsable informatique de l'entreprise JC Decaux.

L'open data, c'est surtout un travail sur l'architecture de notre système d'information. Techniquement, il s'agit d'une duplication de la chaîne de donnée, avec une architecture parallèle : une chaîne dédiée à l'exploitation, l'autre à la diffusion des données. Cela a nécessité un petit investissement pour mettre en place la plateforme dans le cloud et l'architecture. [...] Dans ma logique, l'enjeu était de se dire, on n'est pas capable de protéger nos données, tout le monde vient les extraire sur nos sites, autant leur donner une porte d'accès « propre » pour protéger nos « frontaux » (sites)².

L'infrastructure de diffusion du Grand Lyon est basée sur une architecture similaire à celle décrite ici. Les données sont extraites de la base de données de production (back-office), puis dupliquées dans une base de données intermédiaire (middle office) sur laquelle sont effectués des traitements, avant d'être dupliquées à nouveau sur une base de diffusion (front-office)³. Cette architecture permet aux producteurs de conserver toute maîtrise sur la donnée initiale. Ils restent propriétaires de la donnée et peuvent la modifier en toute autonomie. Le

¹ Les données sont dupliquées dans une nouvelle infrastructure. En isolant les infrastructures de production et de diffusion, cette indépendance renforce la sécurité de la première.

² Source : Entretien Thomas, direction des systèmes d'information, JC Decaux (27 août 2014)

³ Des développements informatiques spécifiques sont sous-traités par le Grand Lyon à un prestataire avec donc un coût financier de mise en œuvre.

système de duplication multiple assure également la sécurité des bases de données initiales, modifiables uniquement par le producteur, et la robustesse de l'infrastructure de diffusion en cas d'usage massif des données par des utilisateurs externes. Dans l'architecture de l'infrastructure de diffusion sont ainsi pris en compte des utilisateurs imaginés : malveillants en rendant difficile le piratage des bases de données ; nombreux dans les cas d'accès massif et simultané aux données ouvertes.

Ce « moissonnage » automatique des données d'une base de production à une autre s'est heurté néanmoins à des limites techniques qui ont contraint la diffusion de certaines données. En effet, les données volumineuses ne parviennent pas à transiter via ces flux. « *Les données 3D posent un problème technique de mise à disposition : leur poids est beaucoup trop important, elles ne "passent pas dans les tuyaux".* »¹ Certaines données ne parviennent pas à passer l'épreuve de l'extraction. Pour dépasser ces data frictions, un découpage communal de ces données a été effectué réduisant le poids des fichiers.

Encadré 10 : L'impossible extraction des données

L'extraction de la donnée de son infrastructure de production n'est pas toujours possible notamment lorsque l'infrastructure est ancienne ou que les données sont insérées dans des systèmes propriétaires. Si le cas ne s'est pas présenté au sein de la communauté urbaine de Lyon, l'autorité organisatrice des transports de Philadelphie, la SEPTA, est confrontée à ce phénomène. Elle est propriétaire des données de localisation en temps réel des trains, mais ces données sont « enfermées » dans des systèmes propriétaires, et les entreprises les ayant mis en place réclament des montants très élevés pour effectuer les développements permettant la mise à disposition des données. « *La limite que l'on a, c'est avec les systèmes propriétaires. Les mecs, ils te demandent des sommes astronomiques, je crois que pour un système il nous demandait quelque chose comme un million de dollars pour ouvrir la donnée. Alors que les bases de données sont juste là [dans le data center situé dans le bâtiment de SEPTA], on a juste à mettre en place un connecteur pour dupliquer la donnée, mais on est bloqué par des contrats qui ont été signés il y a vingt ans quand personne n'imaginait que l'on aurait des smartphones et besoin de données en temps réel².* »

¹ Source : Journal de terrain (10 septembre 2013)

² Source : Entretien Camille, chef de projet informatique, SEPTA, Philadelphie (5 décembre 2014)

*

* *

Contrairement à l'extraction manuelle, l'extraction automatique parvient à résoudre la tension entre le détachement des données de leur infrastructure initiale et leur attachement à de nouveaux utilisateurs. Cette tension est résolue par une duplication de la chaîne de la donnée à l'issue de sa production : l'une est destinée à l'utilisation interne, l'autre à la diffusion ; mais la donnée est similaire pour ces deux utilisations. Par la mise en place de chaîne d'extraction, les données sont détachées de leur infrastructure interne sans être dissociées totalement de la base de production. Ce travail de « dédoublement » crée un second réseau technique parallèle. Une nouvelle logique de gestion de la donnée apparaît alors, avec l'émergence de nouveaux êtres, les « données ouvertes » qui poursuivent une nouvelle trajectoire au sein d'un environnement externe sans rompre le lien avec les entités initiales. La conservation de cette association est cruciale pour assurer la solidité des données ouvertes.

On veut s'assurer que l'information soit fiable et garantie, et ne pas porter la responsabilité si les données réutilisées ne sont pas à jour et qu'un mec loupe son avion à cause de cela. Aujourd'hui, on fournit nos données à la SNCF et à Rhône-Express [l'opérateur assurant la liaison entre Lyon et l'aéroport]. Chez Rhône Express, ils affichaient la donnée dans leurs navettes, mais ils avaient un problème de connectivité avec notre base, et ils affichaient les infos de départ des avions de la veille. Donc, les usagers étaient stressés parce qu'ils ne voyaient pas leur avion. Depuis, ils ont arrêté de diffuser l'info¹.

Cet exemple des données d'information sur les arrivées et les départs de l'aéroport de Lyon révèle l'importance de maintenir un attachement entre les bases de données de production et de diffusion. Le manque de fiabilité des connecteurs empêche l'association durable des données à des utilisateurs externes.

B. Travailler la donnée : le « nettoyage »

Une fois extraites, les données ne sont pas diffusées telle qu'elles. Elles subissent différentes opérations qui visent à rendre possible le détachement de leur environnement de production et leur attachement à de nouveaux utilisateurs. Ces opérations de « brutification² » permettent de rendre la donnée « autonome » de tout environnement social. Ces opérations de « nettoyage » de la donnée peuvent être manuelles ou automatisées. Dans le cas du Grand

¹ Source : Journal de terrain (24 mars 2014)

² J. Denis et S. Goëta, « La fabrique des données brutes. Le travail en coulisses de l'open data. », art cit.

Lyon, elles sont automatisées avec la mise en place d'outils de traitement ETL, qui, outre l'extraction, réalisent les tâches de nettoyage des données préalables à leur diffusion.

Avant sa diffusion, la donnée subit une nouvelle épreuve qui consiste à renommer ou retirer certains de ses attributs afin de supprimer les références à son environnement de production pour permettre son insertion au sein d'une nouvelle infrastructure informationnelle.

La suppression des attributs « métiers » participe au processus de standardisation nécessaire à toute diffusion de données. Chaque administrateur métier choisit, à sa discrétion, les attributs qu'il juge essentiels ou intéressants pour un utilisateur imaginé, en retirant l'ensemble des attributs spécifiques à l'utilisation initiale par les services.

J'ai fait un tour de tous les attributs en disant : « cet attribut peut être intéressant, parce que c'est localisant, c'est une adresse, c'est les dates. » J'essayais toujours de garder les dates de mise à jour dans les attributs, des noms, enfin, des choses qui sont compréhensibles quoi. Donc là, c'est un tri qui était complètement subjectif que j'ai fait de tous les attributs, en essayant aussi de pas trop en mettre¹.

La sélection des attributs s'effectue en premier lieu par la suppression de toutes les informations jugées utiles uniquement pour l'usage interne de la donnée. Cette sélection s'effectue en préfigurant un usage de la donnée ouverte, ici par un citoyen souhaitant connaître les jours des marchés.

Enquêteur : Et sur la mise à disposition de la donnée, tu disais sélectionner certains attributs, ça on prend, ça on prend pas pour diffuser. Il y a certains attributs que tu invisibilises en fait ?

Pauline : Oui, oui. Par exemple, pour les marchés alimentaires, t'avais vraiment énormément de champs à remplir parce que tu disais par exemple, ce marché, imaginons, il a lieu le lundi matin uniquement, tu avais lundi, mardi, mercredi, tous les jours de la semaine avec oui, non, non, non. Après, est-ce qu'il est désinfecté le lundi matin ? Donc, lundi, oui, non, non, non, non. Tu vois à de multiples reprises. Est-ce qu'il est lavé ? Etc. Donc, c'était énorme, tu n'allais pas mettre tous ces champs, c'était ridicule. Même savoir s'il était désinfecté, lavé, c'était vraiment de la donnée métier, donc ça je l'ai pas mis. J'ai mis, ouverture, début du marché, heure de début, heure de fin de marché, et, c'est tout. **En gros, j'ai essayé de me mettre dans la place d'un usager.** Je me suis dit, voilà, il s'en fout de la surface, il s'en fout de la mise en circulation après, parce qu'il y a un temps entre la fin du marché et le début de la mise en circulation où tu peux nettoyer. Donc, voilà, j'ai essayé de la simplifier, de ne mettre que quelques données qui pouvaient être intéressantes pour un usager.

Enquêteur : Et donc, toi tu te mettais, tu t'imaginais un service qui aurait pu être créé et à partir de là, des informations dont il aurait besoin.

¹ Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

Pauline : Moi, c'est vrai qu'à l'époque, c'était plus un usager qui lisait la donnée, qui lisait l'attribut. Alors que c'est pas le cas en fait. Mais je ne me mettais pas dans la peau d'une personne réutilisateur qui ferait un service à partir de ça. Je m'imaginai plus à la place d'un usager qui lit la donnée et qui a envie de savoir quand le marché commence, quand il se termine, quel jour et si c'est plus de l'alimentaire ou du forain¹.

Les usages projetés influencent les caractéristiques de la donnée publiée. C'est également le cas lorsque ce sont des attributs considérés comme « sensibles » qui sont retirés. Par exemple, sur demande du service des arbres, l'attribut indiquant leur état phytosanitaire a été retiré de la donnée « arbre d'alignement ». Il considère que cette information « métier » nécessite « *d'être un professionnel de la botanique, de l'élagage pour analyser ces choses là* »². L'entreprise JC Decaux retire des données publiées l'ensemble des informations qui permettraient d'accéder aux « savoir-faire » de l'entreprise.

On ne diffuse pas toutes nos données. On s'est focalisé sur les données qui nous semblaient les plus pertinentes pour créer un service à l'utilisateur, et les plus demandées aussi. L'objectif est vraiment le service à l'utilisateur. Notre souhait était également que ces données soient facilement utilisables, qu'elles ne soient pas trop complexes, bref qu'elles soient compréhensibles. On a donc mis à disposition des données pertinentes, mais dans un périmètre restreint, qui ne permet pas de comprendre et d'avoir accès aux savoirs-faires de JCDecaux sur l'exploitation et la maintenance du système et du réseau de VLS. On a également retiré tout ce qui est les données personnelles sur les usagers³.

Les données de transport en commun s'inscrivent également dans le même cas. Ce sont les mêmes données qui sont utilisées pour l'exploitation du réseau, pour l'information des voyageurs et pour la diffusion en *open data*. L'infrastructure doit alors être capable de sélectionner en temps réel les informations de ces données envoyées vers chaque système d'information destiné à un usage particulier.

On paramètre les outils de diffusion pour sélectionner l'info qui est poussée systématiquement à destination du client voyageur. On n'extrait qu'une partie de la donnée brute. On a énormément d'informations dans les données, rien que sur le cœur des informations que j'ai cité, on a des dizaines de tables de données, avec chacune des dizaines d'attributs. [...] Par exemple, chaque fois qu'une ligne de bus va terminer son service, le bus va devoir rentrer à son dépôt, ces éléments là, ils vont être prévus dans les horaires, dans les kilomètres parcourus, etc., parce que c'est du temps passé pour le conducteur, c'est du gasoil qui est brûlé, par contre, ça ne doit pas apparaître au client parce que ça ne correspond à rien du tout. Ces éléments là font partie de la différenciation que l'on peut avoir sur la distinction entre l'utilisation et la diffusion de nos informations⁴.

¹ Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

² Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

³ Source : Entretien Thomas, direction des systèmes d'information, JC Decaux (27 août 2014)

⁴ Source : Entretien Benjamin, chef de projet informatique, Keolis Lyon (13 août 2014)

Ce retrait des indications métiers est manifeste également par la modification des noms des champs des attributs afin de faciliter leur compréhension par des utilisateurs externes. « 99% des personnes doivent comprendre, et les listes de valeur, on les décode, comme ça, il n'y a que du texte, et des valeurs intelligibles¹. »

On transforme aussi la donnée pour faire en sorte qu'elle soit intelligible par le réutilisateur final. Parfois, on a un champ interne ou un champ chez le prestataire, qui s'appelle ab413c, il faut peut être le traduire comme étant le nom, ou un champ libellé. Je caricature parce que c'est jamais aussi incompréhensible que ça mais on essaye de rendre intelligible certaines données, pour qu'un utilisateur, s'il a pas un mode d'emploi trop complexe puisse savoir de quoi il s'agit².

Par exemple, pour la donnée « Toilettes publiques », tous les titres, ainsi que certains champs ont été renommés. Les codes postaux des communes ont été remplacés par le nom de la commune, tout comme le numéro identifiant la subdivision responsable du nettoyage. On peut également remarquer le nombre d'attributs métiers qui ont été retiré. De manière similaire, près de 70 attributs sont retirés de la donnée indiquant les arrivées à l'aéroport Saint-Exupéry, et une indication textuelle précise le contenu de chacun des champs.

¹ Source : Entretien Matthias, administrateur de données, Grand Lyon (17 janvier 2014)

² Source : Entretien François, chef de projet *open data*, Grand Lyon (11 février 2015)

Illustration 24 : Flux des arrivées de l'aéroport Lyon Saint-Exupéry

Liste des informations disponibles pour chacun des vols arrivant à l'aéroport
Les attributs barrés sont ceux qui ne sont pas disponibles en *open data*

```
<GetArrivalsOUT> Racine du body de la réponse SOAP
<MetaData> Metadonnées de la réponse
  <User> identifiant appelant
  <RequestDate> date d'appel
  <ResponseDate> date de fin de préparation du flux de réponse (avant transport)
  <Parameters>
    <MinDate> paramètre MinDate de l'appel
    <MaxDate> paramètre MaxDate de l'appel
  </Parameters>
  <ResultSetSize> Nombre de vols ramenés dans le résultat
</MetaData>
<Arrivals> Liste des vols arrivée
  <Arrival> Vol arrivée - Élément répétable pour chaque vol arrivée dans le résultat
    <InternalId> Identifiant technique du vol / unique / issu d'une séquence
    <FlightStatus> Etat du vol
      <Code> Code Etat du vol
      <Description> Description Etat du vol
    </FlightStatus>
    <Remark> Remarque (irrégularité)
      <Code> Code Remarque
      <Description> Description Remarque
    </Remark>
    <FlightNumbers>
      <InternalFlightNumber> Numero Vol Interne SIA (exploitant commercial)
      <TripNumber> Numero de Ligne (exploitant commercial)
      <Suffix> Suffixe (exploitant commercial)
      <IATAFlightNumber> Numero Vol format IATA (code IATA compagnie + espace + n° ligne) (exploitant commercial)
      <CAOFlightNumber> Numero Vol format OACI (code OACI compagnie + n° ligne + suffixe) (exploitant commercial)
      <OperatorInternalFlightNumber> Numero Vol Interne SIA (opérateur)
      <OperatorTripNumber> Numero de Ligne (opérateur)
      <OperatorSuffix> Suffixe (opérateur)
      <OperatorIATAFlightNumber> Numero Vol format IATA (code IATA compagnie + espace + n° ligne) (opérateur)
      <OperatorCAOFlightNumber> Numero Vol format OACI (code OACI compagnie + n° ligne + suffixe) (opérateur)
      <Callsign> Callsign (identifiant vol pour le contrôle aérien)
    </FlightNumbers>
    <Timestamps> Horodatages sur le vol
      <SIBT> Horaire planifié (commercial)
      <ATOT_OUTSTATION> Heure de décollage de l'aéroport de provenance
      <ELDT> Heure estimée d'atterrissage
      <EIBT> Heure estimée d'arrivée au point de stationnement
      <ALDT> Heure constatée d'atterrissage
      <AIBT> Heure constatée d'arrivée au point de stationnement
      <CancellationDate> Date d'Annulation du vol
      <DiversionDate> Date de Deroutement du vol
      <ModificationDate> Date de dernière modification du vol dans le SIA
    </Timestamps>
    <Aircraft> Appareil
      <Registration> Immatriculation
      <AircraftType> Type Appareil
        <IATACode> Code IATA Type Appareil
        <CAOCCode> Code OACI Type Appareil
        <ModelName> Nom de Modele Type Appareil
      </AircraftType>
    </Aircraft>
    <AirportResources> Ressources aéroportuaires allouées au vol
      <Terminal> Terminal
      <Runway> Piste
      <Stand> Stationnement
        <BlockStand> Parking sur lequel le vol fait block
        <StandsList> Liste des parkings (détail)
          <Stand> Parking - élément répétable pour chaque parking si tractage(s)
            <InternalId> Identifiant technique du vol / unique / issu d'une séquence
            <Code> Code ressource
            <ScheduledBegin> Date Debut Planifiée d'Usage
            <ScheduledEnd> Date Fin Planifiée d'Usage
            <ActualBegin> Date Debut Réalisée d'Usage
            <ActualEnd> Date Fin Réalisée d'Usage
            <Order> Numero d'Ordre d'Usage
            <ModificationDate> Date de dernière Modification d'Usage
          </Stand>
        </StandsList>
      </Stand>
    </AirportResources>
    <Unboarding> Débarquement
      <Airbridges> Passerelles - liste de codes de ressources séparés par ','
      <Gates> Portes - liste de codes de ressources séparés par ','
      <GatesList> Liste des Portes (détail)
        <Gate> Porte - élément répétable pour chaque porte allouée au vol
          <InternalId> Identifiant technique du vol / unique / issu d'une séquence
          <Code> Code ressource
          <ScheduledBegin> Date Debut Planifiée d'Usage
          <ScheduledEnd> Date Fin Planifiée d'Usage
          <ActualBegin> Date Debut Réalisée d'Usage
          <ActualEnd> Date Fin Réalisée d'Usage
          <Order> Numero d'Ordre d'Usage
        </Gate>
      </GatesList>
    </Unboarding>
  </Arrival>
</Arrivals>
```

```

        <ModificationDate> Date de dernière Modification d'Usage
    </Gate>
</GatesList>
</Unboarding>
<BaggageDelivery> Livraison baggages
    <ArrivalLounge> Salle d'Arrivée
    <BaggageBelts> Convoyeurs - liste de codes de ressources séparés par ','
    <Carrousel> Carrousel - liste de codes de ressources séparés par ','
    <BaggageBeltsList> Liste des convoyeurs (détail)
        <BaggageBelt> Convoyeur - élément répétable pour chaque convoyeur alloué au vol
            <InternalId> Identifiant technique du vol / unique / issu d'une séquence
            <Code> Code ressource
            <ScheduledBegin> Date Debut Planifiée d'Usage
            <ScheduledEnd> Date Fin Planifiée d'Usage
            <ActualBegin> Date Debut Réalisée d'Usage
            <ActualEnd> Date Fin Réalisée d'Usage
            <Order> Numéro d'Ordre d'Usage
            <ModificationDate> Date de dernière Modification d'Usage
        </BaggageBelt>
    </BaggageBeltsList>
</BaggageDelivery>
</AirportResources>
<GroundHandlingList> Liste des assistances sur le vol
    <Handling> Assistance
        <HandlerCode> Code Assistant
        <HandlerName> Nom Assistant
        <HandlingType> Type Assistance
        <ModificationDate> Date Modification
    </Handling>
</GroundHandlingList>
</Airlines> Compagnies aériennes
    <Airline> Compagnie exploitant commercial
        <IATACode> Code IATA compagnie
        <CAOCCode> Code OACI compagnie
        <Name> Nom compagnie
    </Airline>
    <Operator> Compagnie opérateur
        <IATACode> Code IATA compagnie
        <CAOCCode> Code OACI compagnie
        <Name> Nom compagnie
    </Operator>
</Airlines>
</Airports> Aéroports de provenance
    <Origin> Aéroport d'origine (1er aéroport de la route du plan de vol)
        <IATACode> Code IATA Aéroport
        <CAOCCode> Code OACI Aéroport
        <Name> Nom Aéroport
    </Origin>
    <Previous> Aéroport précédent (dernière escale avant LYS de la route du plan de vol)
        <IATACode> Code IATA Aéroport
        <CAOCCode> Code OACI Aéroport
        <Name> Nom Aéroport
    </Previous>
</Airports>
<CountryType> Type de pays de l'escale d'origine (schengen, U/E non schengen, international)
    <Code> Code
    <Description> Description
</CountryType>
<ServiceType> Type de service IATA (régulier, charter, etc.)
    <IATACode> Code IATA
    <Description> Description
</ServiceType>
<OperationalComment> Commentaire pour les Opérations ADL
<PublicComment> Commentaire Publique
<CodeSharesList> Liste des Code shares
    <CodeShare> Code Share - Élément répétable pour chaque code share du vol
        <InternalId> Identifiant technique du vol / unique / issu d'une séquence
        <InternalFlightNumber> Numéro de Vol Interne
        <TripNumber> Numéro de Ligne
        <IATAFlightNumber> Numéro de Vol format IATA
        <CAOFlightNumber> Numéro Vol format OACI
        <Airline> Compagnie
            <IATACode> Code IATA
            <CAOCCode> Code OACI
            <Name> Nom
        </Airline>
        <ModificationDate> Date de Modification
    </CodeShare>
</CodeSharesList>
<DelaysList> Liste des DL (codes retards)
    <Delay> DL - élément répétable pour chaque DL associé au vol
        <InternalId> Identifiant technique du vol / unique / issu d'une séquence
        <IATACode> Code Numérique DL
        <AlphaCode> Code Alpha DL
        <Description> Description
        <Duration> Durée
        <ModificationDate> Date de Modification
    </Delay>
</DelaysList>
<TurnFlightInternalId> Identifiant technique du vol / unique / issu d'une séquence du vol départ en rotation
</Arrival>

```

Comparaison de la donnée métier et de la donnée diffusée (Extraits de la donnée « Toilettes publics »)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
OBJECTID	DTH_CREATI	DTH_DTH_MAJALP	SID_R_SID_EEMPLAC	STR_NUMVOI	STR_NUMERO	STR_CODEFU	STR_OBSLOC	SID_COMMUN	STR_COMMENT	ICD_ETAT	ICD_GESTIO	INT_NBWC	INT_NBURIN	INT_NBLAVA	ICD_ENTREP	ICD_LUNDI	ICD_MARDI	ICD_MERCRE	ICD_JEUDI	ICD_VENDRE	ICD_SAMEDI	ICD_DIMANC	FLT_FREQUE		
3	20/04/06	20/04/06	W12301	Place du 11 Novembre 1918 et	76	20589	Fae n76.	69260		1	5	1	1	1	5	1	1	1	1	1	1	1	0 6.00		
32	17/12/09	17/12/09	W45508	Rue du Cimetire			0 Face au cimetire.	69191		1	4	1	2	0	5	1	1	1	1	1	1	1	0 6.00		
33	17/12/09	17/12/09	W43618	Route du Mont Cindre	N80		0	69191	Ferm durant l' hive	1	4	1	0	0	5	1	1	1	1	1	1	1	0 6.00		
41	24/07/12	24/07/12	W85552	Chemin de la Chapelle			24914 Parking de la chapelle vers Route de	69286		1	4	1	0	1	7	1	1	1	1	1	1	1	0 6.00		
57	08/02/12	08/02/12	W78378	Place de la Mairie			30113 Sous le parking couvert, au bas escc	69278		1	4	2	0	2	7	1	1	1	1	1	1	1	0 6.00		
60	16/03/06	16/03/06	W99556	Place de l'Europe	4		8541 face la bibliotheq	69100		1	5	2	3	0	5	2	2	2	2	2	2	2	0 12.00		
61	13/04/06	18/04/06	W70776	Place Bad Abbach			30016 sous la mairie	69044		1	5	1	1	0	5	1	1	1	1	1	1	1	0 6.00		
65	13/04/06	13/04/06	W45570	Place Charles de Gaulle			6810 face au numiro 6	69081		1	5	1	2	2	5	1	1	1	1	1	1	1	0 6.00		
76	14/04/06	20/04/06	W65186	Chemin de Montray	34		15406 Fae au n 34 (cimetire)	69202		1	5	4	6	0	5	1	1	1	1	1	1	1	0 6.00		
77	14/04/06	20/04/06	W54174	Place Henri Rvy	93		15489 Fae n 93 route de la Libration.	69202		1	5	2	0	0	5	1	1	1	1	1	1	1	0 6.00		
79	14/04/06	20/04/06	W11310	Place de la Paix	6		31393 Fae n 6.	69152		1	5	1	1	0	5	2	2	2	2	2	2	2	0 12.00		
82	14/04/06	20/04/06	W68517	Boulevard de l'Europe			11914 fae rue Vaillant Couturier, square C	69152		1	5	1	2	0	5	1	1	1	1	1	1	1	0 6.00		
86	14/04/06	20/04/06	W62103	Place Kellermann			11621 A proximitt de la rue Elyse Reclus.	69149		1	5	2	3	0	5	1	1	1	1	1	1	1	0 6.00		
89	14/04/06	20/04/06	W57835	Rue Pierre Smard			11715 Angle rue Aulagne	69149		1	5	0	3	0	5	1	1	1	1	1	1	1	0 6.00		
90	14/04/06	20/04/06	W16491	Rue Etienne Dolet	1		11677 A ct du n 1.	69149		1	5	1	0	1	5	1	1	1	1	1	1	1	0 6.00		
95	14/04/06	20/04/06	W44624	Chemin de Presles			8456 Intrieur du nouveau cimetire.	69100		1	5	2	0	1	5	1	1	1	1	1	1	1	0 6.00		
99	14/04/06	14/04/06	W57545	Place de la Porte de Lyon			31899	69072		1	5	1	2	1	5	1	1	1	1	1	1	1	0 6.00		
101	14/04/06	14/04/06	W33315	Rue de l'Eglise			15926 place de l'glise droite de l'glise	69205		1	5	1	2	1	5	1	1	1	1	1	1	1	0 6.00		
103	18/04/06	18/04/06	W65533	Rue de la Chapelle de Bel Air	2		7381	69089		1	5	1	2	1	5	1	1	1	1	1	1	1	0 6.00		
104	19/04/06	19/04/06	W77598	Place Andre-Marie Perrin	82		5846	69069		1	5	1	0	1	5	2	2	2	2	2	2	2	0 12.00		

Illustration 25 : La donnée « Toilettes publics » extraite de la base métier

A	B	C	D	E	F	G	H	I	J
commune	voie	numerodans	gestionnai	observatio	miseajoura	miseajourg	gid		
TASSIN-LA-DEMI-LUNE	Rue Professeur Deperet	81	NET Sud-Ouest	sur le parking du rond point d'Ala	19/04/06		1		
TASSIN-LA-DEMI-LUNE	Avenue du 8 Mai 1945	81	NET Sud-Ouest		19/04/06		2		
VERNAISON	Place du 11 Novembre 1918 et du 8 Mai 1945	76	NET Sud-Ouest	Fae n76.	20/04/06		3		
TASSIN-LA-DEMI-LUNE	Rue de Belgique		NET Sud-Ouest		03/05/06		4		
FRANCHEVILLE	Rue de la Doulline	2	NET Sud-Ouest		26/01/07		5		
SAINT-CYR-AU-MONT-D'OR	Rue du Cimetire		NET Nord-Ouest	Face au cimetire.	17/12/09		6		
SAINT-CYR-AU-MONT-D'OR	Route du Mont Cindre	N80	NET Nord-Ouest		17/12/09		7		
SAINT-CYR-AU-MONT-D'OR	Place de la Rpublique		NET Nord-Ouest		17/12/09		8		
NEUVILLE-SUR-SAONE	Avenue Jean Christophe		NET Nord-Ouest	Place Jean Christophe	25/01/12		9		
NEUVILLE-SUR-SAONE	Rue Pierre Dugelay		NET Nord-Ouest	Angle Avenue du 11 Novembre 1918.	25/01/12		10		
NEUVILLE-SUR-SAONE	Rue Rey Loras		NET Nord-Ouest	Parking Place de Verdun.	25/01/12		11		
NEUVILLE-SUR-SAONE	Place du March		NET Nord-Ouest	Vers Pont de Neuville.	25/01/12		12		
RILLIEUX-LA-PAPE	Rue de la Barse		NET Nord-Ouest	Vers Place du Chteau.	24/07/12		13		
MONTANAY	Rue Centrale	Face N731	NET Nord-Ouest	Vers Place de la Poype.	08/02/12		14		
RILLIEUX-LA-PAPE	Chemin de la Chapelle		NET Nord-Ouest	Parking de la chapelle vers Route de Strasbourg.	24/07/12		15		
ALBIGNY-SUR-SAONE	Rue Etienne Richerand		NET Nord-Ouest		13/06/12		16		
ALBIGNY-SUR-SAONE	Place du Docteur Aloin		NET Nord-Ouest	Derrire glise.	13/06/12		17		
GENAY	Rue des Ecoles	Face N64	NET Nord-Ouest	Dans le Square Gnral de Gaulle.	08/02/12		18		
SATHONAY-CAMP	Boulevard Castellane	Face N43	NET Nord-Ouest		23/07/12		19		
SATHONAY-CAMP	Rue de la Poste		NET Nord-Ouest		23/07/12		20		
RILLIEUX-LA-PAPE	Place Marchal Juin		NET Nord-Ouest		24/07/12		21		
ALBIGNY-SUR-SAONE	Quai Gnral de Gaulle		NET Nord-Ouest	Halte Fluviale.	13/06/12		22		

Illustration 26 : La donnée « Toilettes publics » disponible sur SmartData

Sélectionner et renommer les attributs contribue à la fabrication d'une donnée « brute », détachée de son environnement initial pour pouvoir être utilisée par d'autres acteurs. Toutefois, l'équilibre entre détachement du producteur et attachement aux réutilisateurs est complexe. La suppression de ces attributs peut considérablement réduire la richesse initiale de la donnée et son sens, ce qui la rend inutilisable pour un acteur externe. Par exemple, ce producteur estime que les choix effectués avant son arrivée rendent peu exploitable la donnée des circuits de viabilité hivernale.

Enquêteur : Oui, parce que du coup, il y a une sélection de ce qui est publiable ?

Oui, ce qui est publiable ou pas. Enfin... Par exemple, si je vais sur le déneigement... [...] En gros, sur SmartData, on a le nom de la voie, date de création du tronçon, date de mise à jour alphanumérique, date de mise à jour graphique, je suis pas sûr que ça serve ça par exemple sur SmartData, on a le nom de la voie, et la longueur. C'est tout. Sur SmartData, on ne diffuse pas la priorité du circuit, si c'est une priorité 1, 2 ou 3, de nuit, des choses comme ça. Et on ne diffuse pas le traitement. Si on passe au centre, à droite ou à gauche. Il y a peu d'attributs, franchement, la personne qui veut exploiter les données déneigement dans SmartData, pfff... C'est pas simple... [...] Limite je trouve ça bête que les circuits de VH y soient de manières altérées. Limite, je préférerais qu'il n'y en ait pas. Plutôt que d'avoir juste le circuit, et qu'au final il n'y a aucune info, c'est juste inutilisable. Ça ne sert pas à grand chose.

Enquêteur : T'as l'impression que la qualité est un peu dégradée ?

Oui. Là, on a juste voulu dire : « *on met quelque chose dans SmartData pour dire de mettre quelque chose* ». Mais, au final, est-ce qu'il n'y aurait pas fallu mettre, peut-être pas tous, mais on met toutes les premières priorités, les lignes fortes, ce que l'on traite vraiment tout de suite. Et qui peut servir à dire, voilà, je sors de chez moi, je sais qu'ils passent dans cette rue, je suis sûr de pas avoir de neige. Parce que c'est une ligne prioritaire au Grand Lyon, elle sera déneigée. Tandis que là, j'ai tout... Personne ne le sait, mais, oui, il y a les hauts le pieds, y'a l'acheminement, y'a le dépôt de sel. C'est des tronçons qui ne sont pas déneigés. Je pense qu'il aurait fallu plus sélectionner certaines données. Peut-être pas mettre tous les circuits mais ceux qui ont un intérêt. En gros, je pense qu'ils ont été cachés parce qu'il y a une troisième priorité, et, on ne les déneige quasiment jamais parce qu'on a pas le temps d'aller dessus. Mais, il aurait juste fallu les masquer. Et on montre là où on passe, où on est sûr de passer et d'assurer la prestation. Après, là où on n'est pas sûr d'assurer la prestation, on le met pas. Là, on sait que l'on passe partout, mais c'est pas la réalité quoi ! (rires)¹

Cette perte de richesse rend la donnée peu exploitable pour un éventuel utilisateur car la représentation des circuits de viabilité hivernale qu'elle transmet ne correspond au traitement effectué par les services du Grand Lyon. Volontairement, les producteurs peuvent choisir de ne pas publier certaines informations afin de restreindre les réutilisations possibles. Comme le sous-entend cet interlocuteur, la priorité de déneigement n'est pas indiquée pour ne pas rendre visible le fait que certaines voies ne sont jamais déneigées. Le travail de « brutification » rend

¹ Source : Entretien, Louis, administrateur de données, Grand Lyon (22 novembre 2013)

ici inopérable la donnée ; dès lors, si elle est détachée de son environnement initial, son association à un utilisateur externe est compromise.

*

* *

Ces opérations de « nettoyage » de la donnée révèlent que la donnée « brute » est un mythe. Il n'existe pas de données autonomes de tout environnement social. La donnée est toujours travaillée pour répondre à un usage déterminé et/ou empêcher un usage préfiguré. Il n'y a ainsi pas de rupture entre données et usages comme cela est trop souvent sous-entendue dans les discours autonomisant les données de tout environnement social.

C. Diffuser la donnée : plateforme, formats et métadonnées comme « opérateurs de circulation »

Une fois les données prêtes à être diffusées, il reste à assurer la circulation des données de l'infrastructure du Grand Lyon à celle d'utilisateurs externes. Pour que les données puissent être utilisées dans un nouvel environnement, différentes médiations leur sont associées : une infrastructure de diffusion, un format et des métadonnées. Ces opérateurs de circulation favorisent leur intégration dans d'autres systèmes techniques en permettant le détachement des données de leurs infrastructures initiales.

1) L'infrastructure de diffusion

L'infrastructure de diffusion de données est la plateforme informatique assurant la mise en relation entre les systèmes d'information des producteurs et des réutilisateurs. Ce positionnement d'intermédiaire fait de cette infrastructure un instrument crucial pour réguler les flux de données. Or, le Grand Lyon a souhaité « *ne pas perdre la main, garder la maîtrise dessus, car la plateforme est au cœur de tout*¹. » Cet aspect a été central dans le choix du Grand Lyon de développer sa propre plateforme, basée sur des logiciels libres, plutôt que d'avoir recours à des solutions commerciales. En effet, des entreprises, notamment IBM, ont proposé à la communauté urbaine de lui fournir, gratuitement, dans le cadre d'un programme de R&D, une plateforme de services autour de la donnée. « *Ils veulent nous vendre leur*

¹ Source : Entretien Nicolas, direction des systèmes d'information, Grand Lyon (16 janvier 2013)

solution globale de plateforme de services. Mais, le risque est que l'on soit ferrailé avec eux, que l'on se retrouve pieds et poings liés »¹. L'enjeu de lock-in est d'autant plus important qu'ils ne concernent pas uniquement la communauté urbaine de Lyon, mais également l'ensemble des entreprises développant des services à partir des données ouvertes sur cette plateforme. Pour éviter cette irréversibilité et garantir la pérennité des modalités d'accès aux données, le Grand Lyon a fait le choix d'utiliser une solution d'infrastructure de diffusion de données géographiques, basée sur des logiciels libres.

2) Le « formatage » : des formats ouverts pour favoriser l'usage large des données

Déterminant le champ de l'usage possible des données, le format est un enjeu crucial de la politique de diffusion. Les promoteurs de l'*open data* militent pour une diffusion des données sous des formats ouverts et interopérables afin de permettre leur utilisation la plus large possible². Le Grand Lyon propose les données sous différents formats à la fois en flux et en téléchargement. Le choix des formats varie selon le type de données (géographiques, transport, alphanumériques, etc.). Néanmoins, des formats ouverts et standardisés sont privilégiés afin de favoriser la plus grande interopérabilité. Le chef de projet revient avec nous sur ces choix de format.

Sur les données alphanumériques, on est sur des logiques de CSV³, de choses assez simple à lire par un humain. Pour le coup, c'est pas une standardisation, une normalisation, c'est plus un espèce de standard de fait, standard d'usage⁴.

S'il n'est pas un standard normalisé, le choix du CSV repose sur sa simplicité d'utilisation, aisément compréhensible à la fois par des systèmes informatiques et des « humains ». Ces critères sont similaires pour le choix de la diffusion des données de transport en commun au format GTFS.

¹ Source : Journal de terrain, réunion avec Karine Dognin-Sauze (15 avril 2013)

² Les militants de l'*open data* juge du degré d'ouverture d'une institution en fonction des formats de mise à disposition de données. Cf. Samuel Goëta, « The Daily Shaping of State Transparency: Emerging Standards in Open Government Data », Buenos Aires, 2014.

³ Le *Comma-separated values* est un format ouvert de représentation textuelle de données tabulaires, chaque valeur étant séparée par des virgules.

⁴ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

Pour les horaires théoriques [des transports en commun], on a voulu mettre un format qui soit le plus standard possible dans l'utilisation, donc, on est parti sur du GTFS¹, ce standard de fait, qui est depuis quelques années fortement utilisé par les communautés de ceux qui travaillent sur les offres de transport. [...] On s'inscrit dans une logique de dire, plus on va respecter un standard, plus ce sera facilitateur pour l'utilisation. C'est l'objectif de l'open data : on veut que l'information soit utilisée au maximum. Si on met un format spécifique, forcément, on va avoir des consommateurs qui vont avoir des réticences à l'utiliser, qui vont nous solliciter en disant : « *on comprend pas, il manque si, il manque ça, etc.* ». Là, on a des choses pour lesquelles on n'a pas forcément d'explications à donner. Vous avez les données de l'offre de transport du réseau TCL, dans le format GTFS, si la personne ne connaît pas le sujet, elle ira voir à quoi correspond le format GTFS. Si le consommateur a déjà développé des petits outils, des applis, où il a des connecteurs standard, il a plus qu'à dire : « *super je prends les données de TCL, je les met là* », et finalement en cinq minutes, j'ai mon calculateur dédié PMR, ou je ne sais pas quoi qui est conçu².

Le choix de formats standardisés détermine également les modalités de diffusion aux données géographiques, accessibles à la fois en téléchargement et par des flux automatisés, afin de permettre l'association d'une pluralité d'acteurs aux données ouvertes.

Il y a eu un choix de s'appuyer sur la normalisation OGC pour les données cartographiques notamment. Ce choix a été assez structurant. C'est un choix pertinent. Aujourd'hui, la plupart des SIG savent interroger ça, INSPIRE demande ça, les outils de webmapping type google, type openlayers et autres savent intégrer ces systèmes normalisés³.

Les données géographiques sont également disponibles en téléchargement au format *shapefile*, un format très largement utilisé par les différents logiciels d'information géographique. « *le shape est un format ouvert, un format standard maintenant, même s'il était propriétaire au début. Tous les logiciels SIG savent manger du shape*⁴. » Toutefois, ces formats standardisés sont spécifiques au milieu de la géomatique. Pour élargir l'usage des données à d'autres publics, les données géographiques sont également mises à disposition au format KML⁵, un langage développé initialement pour l'utilisation du logiciel *Google Earth*, mais qui est devenu un standard international.

Ce choix du KML, il a été fait dans une logique de dire aujourd'hui toutes les données qui sont publiées sur la plateforme elles sont publiées qu'avec des formats très techniques, peu digestes

¹ Le *General Transit Feed Specification* est un format standardisé spécifique aux données de transport en commun. Il se compose de plusieurs fichiers CSV représentant chacun un élément (points d'arrêt, itinéraires, horaires, etc.).

² Source : Entretien Benjamin, chef de projet informatique, Keolis Lyon (13 août 2014)

³ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

⁴ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

⁵ *Keyhole Markup Language* est un langage destiné à la gestion de l'affichage de données géospatiales.

par les utilisateurs, donc si on veut permettre une compréhension rapide du contenu de la donnée, avoir un aperçu rapide de la donnée, la donnée KML, elle fournit à tout le monde une manière d'accéder à des données avec des logiciels grand public. [...] Je télécharge le KML, j'ouvre mon KML, et je l'ai dans mon Google Earth. Parce que pour les données géographiques, il n'y a que ça qui permet de le faire très simplement¹.

Enfin, pour répondre aux demandes de développeurs informatiques, les données géographiques sont accessibles également via des flux JSON² et GeoJSON³.

Au delà des formats géographiques, on a également volontairement ajouté une mise à disposition sous un format non géographique, le JSON, qui peut être utilisé pour la création d'applications web ou mobile. On s'est en effet rendu compte que les développeurs web, c'est pas trop leur truc de manipuler du WFS, en tout cas, les développeurs web. Le JSON est un format beaucoup plus intelligible par eux qui ont moins la culture géomatique que ceux qui viennent du monde géomatique pur⁴.

Par le formatage, les données sont calibrées pour répondre aux attentes de certains utilisateurs. Le choix des formats est déterminé par l'usage imaginé des données. Chaque format incorpore en effet une certaine représentation de l'utilisateur préfiguré⁵ : le JSON pour les développeurs, le CSV pour être lisible par des utilisateurs « humains », le shape et les flux WMS/WFS pour les géomaticiens, etc. Pour chacun de ces publics, l'objectif est de lui fournir le format le plus standardisé et le plus ouvert afin de faciliter l'interopérabilité de la donnée, c'est-à-dire de la rendre autonome du système informatique utilisé pour sa production et son usage interne, autorisant ainsi une liberté de choix de logiciel pour exploiter cette donnée. En outre, la pluralité des formats disponibles pour chacune des données permet l'association de publics multiples.

Néanmoins, tout aussi standardisés qu'ils sont, ces formats dissocient certains utilisateurs autant qu'ils en associent d'autres. Ils sont destinés à des publics avertis, des professionnels bien ciblés (développeurs, géomaticiens), et écartent de fait le grand public, mais également des utilisateurs antérieurs de ces données. Les communes ou les prestataires

¹ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

² *JavaScript Object Notation* est un format de données textuelles, qui se caractérise par sa simplicité.

³ *Geographic JavaScript Object Notation* est un format ouvert d'encodage de données géospatiales utilisant la norme JSON.

⁴ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

⁵ S. Goëta, « The Daily Shaping of State Transparency: Emerging Standards in Open Government Data », art cit.

de la communauté urbaine avaient auparavant accès aux données par l'envoi de fichiers aux formats qu'elles désiraient. Maintenant, elles doivent passer par la plateforme pour récupérer ces données, or, comme l'indique une personne du service de l'information géographique : « *les communes ne vont pas passer du jour au lendemain à l'utilisation de flux WMS. Elles ne vont pas changer facilement leurs habitudes, et beaucoup ne sauront d'ailleurs pas faire*¹. » En outre, certains formats ouverts et interopérables ne sont pas toujours les plus répandus et utilisés comme le pointe le producteur des données 3D.

Pour les données 3D, il n'y a pas vraiment de standards. Le format le plus interopérable aujourd'hui, c'est le CityGML, mais personne ne l'utilise. Au quotidien, les gens, ils s'en moquent un peu du CityGML. Quand on ouvre un document word, on aime bien avoir un point doc, on n'aime pas avoir un je sais pas quoi. C'est plus de contraintes pour les utilisateurs. Quand je leur en parle, ça leur passe au dessus. Ça ne les intéresse pas. Ils veulent leur format. S'ils ont un Autodesk, ils voudront du 3DS. Celui qui a un Rhino, il voudra du Rhino, celui qui a SketchUp, il veut du SketchUp².

Publiée dans un format interopérable mais peu utilisé, l'usage des données 3D est plus complexe. En témoignent, les mails d'utilisateurs reçus régulièrement s'interrogeant sur les logiciels permettant d'utiliser ce format qui leur est inconnu.

*

* *

Le formatage consiste également à restructurer l'information contenue dans les bases de données afin que celles-ci soient utilisables pour d'autres modalités que celles prévues initialement. Il s'agit souvent d'opérations simples comme d'associer une adresse à une position géographique.

Dans nos systèmes internes, par exemple, on a des points d'intérêts pour lesquels on ne met pas d'adresse postale, parce qu'on n'en a pas besoin. Mais, dans le cadre de l'open data, on imagine plusieurs usages possibles. L'utilisateur, il peut se baser soit sur une localisation en x;y ; soit sur des adresses postales, s'il veut par exemple rechercher les informations par commune. Donc, on a enrichi un peu le contenu de la donnée pour permettre d'avoir différents types d'utilisation³.

¹ Source : Journal de terrain, Quentin (11 mars 2013)

² Source : Entretien Romain, producteur de données géographiques (21 novembre 2013)

³ Source : Entretien Benjamin, chef de projet informatique, Keolis Lyon (13 août 2014)

Les opérations sont parfois plus complexes. Par exemple, deux sources de données peuvent être croisées pour en produire une troisième jugée plus complète et plus fiable.

Pour la publication des données, on développe un petit script de chargement qui dit simplement voici ma source de données, et voilà où je la charge, et entre les deux, on peut mettre un certains nombres d'étapes de transformation. Si c'est de la donnée géographique, je change la projection, je fais une jointure spatiale avec telle ou telle information. Si c'est de la donnée alpha, je fais une jointure entre deux attributs, je normalise telle et telle chose. [...] Parfois on restructure les données. Je pense aux données trafic, on fait de la fusion entre une donnée de référence qui est le référentiel géographique qui est statique et une donnée temps réel. On transforme aussi un champ de type valeur de comptage, qui évoque peut-être pas grand chose à un certain nombre de personnes, on le traduit en un état vert orange ou rouge, pour que tout de suite il y ait une appréhension possible de la chose. On fait pareil sur les vélov où on va aller récupérer des données sur la disponibilité fournies par JC Decaux et on va les croiser avec notre référentiel spatial de stations que l'on considère de meilleure qualité de précision géographique. Donc on va faire le croisement de ces données là, et on fabrique une troisième donnée qui est en fait une combinaison de deux données différentes¹.

Ces opérations de formatage doivent faciliter l'association de la donnée à des infrastructures techniques variées et répondre à des usages hétérogènes. Elles visent à rendre la donnée la plus « brute » possible. La construction du caractère « brut » de la donnée ne va pas de soi. Elle est le fruit d'allers-retours avec les utilisateurs afin de stabiliser l'identité de la « donnée ouverte » et pérenniser l'association avec les nouveaux utilisateurs. Par exemple, les données de trafic routier sont initialement publiées sous forme de traficolor afin d'assurer une cohérence de l'information fournie. Cependant, s'il garantit une interprétation identique des taux d'occupation de la voirie, ce choix empêche certains usages. Après discussions avec des utilisateurs, le producteur va alors revenir sur cette décision et indiquer les taux de débit et d'occupation dans la structure de la donnée.

3) Les métadonnées

Les métadonnées sont le troisième opérateur de circulation. Constituées de l'ensemble des informations relatives à une donnée, elles visent à faciliter le transfert des données au sein d'un autre univers parfois très éloigné de l'environnement initial. Renseigner des métadonnées, c'est ouvrir la boîte noire de la donnée : expliciter les conventions sur lesquelles elle repose afin de la rendre compréhensible et d'éviter les mésusages.

¹ Source : Entretien François, chef de projet informatique, Grand Lyon (11 février 2015)

L'architecture de diffusion de données de la communauté urbaine de Lyon repose sur un catalogue de métadonnées. Pour chaque donnée diffusée, le producteur doit renseigner plusieurs champs d'une nomenclature standardisant les informations décrivant chacune des données (source, modalités de production, description de la donnée, etc.). Renseigner une fiche de métadonnées est ainsi un point de passage obligé pour l'ensemble des producteurs souhaitant publier une donnée.

Cette injonction à la documentation constitue parfois un frein à la diffusion des données. En effet, tous les agents des collectivités ne disposent pas de l'expertise pour documenter les données qu'ils utilisent quotidiennement, parfois sans en connaître la provenance ou les méthodes de leur production. Seuls quelques professionnels disposent de ces compétences, essentiellement les acteurs dont le métier est de produire et de gérer de la donnée, tels que les administrateurs de données géographiques. En outre, ce travail invisible de documentation fastidieux et chronophage est souvent peu valorisé.

Il fallait absolument que l'on documente les données. On ne pouvait pas sortir aucune donnée tant que les métadonnées n'étaient pas rentrées. Mais, tout le monde a renâclé, personne ne voulait s'y mettre. [...] Ça peut être un frein tout de même¹.

Mais, au final, ça reste quelque chose d'ingrat que de saisir de la métadonnée, qu'elle soit géographique ou pas. Nous, on a considéré que c'était le passage obligé, il n'y a pas de métadonnées, on ne publie pas de données, parce qu'au final, ce qu'on constate, même dans les outils d'open data récents, il y a toujours un peu d'informations descriptives qui accompagnent une donnée, ça c'est incontournable².

Ce travail ingrat est pourtant indispensable à la circulation des données de leur univers de production vers des environnements hétérogènes et favoriser une diversité sociale d'usage. Pour faciliter ce transfert d'un monde à l'autre, certains producteurs ajoutent aux fiches standardisées des descriptions détaillées de leurs données afin de minimiser les risques de trahison³.

¹ Source : Entretien avec Pauline, responsable de la plateforme, ex-administratrice de données géographiques, (13 octobre 2013)

² Source : Entretien avec Etienne, administrateur de données (26 novembre 2013)

³ Ces notices descriptives peuvent être très détaillées. Par exemple, les métadonnées fournies par le Département de planification urbaine de la ville de New York pour sa base de données PLUTO sont longues de plus de 60 pages !

*

* *

D'un contexte initial de forte hétérogénéité, la plateforme participe à la standardisation des données en leur associant une fiche de métadonnées et des formats de téléchargement normalisés. Ces médiations sont autant d' « opérateurs de circulation¹ », qui facilitent le transfert d'un environnement à un autre. En proposant un cadre commun à l'échelle de l'agglomération, ils réduisent l'ambiguïté des données et facilitent la connexion entre les univers sociaux des producteurs et des réutilisateurs.

D. La production des « données ouvertes » : un investissement de forme

Loin d'être disponible et prêtes à l'ouverture, les données subissent un ensemble de traitements avant de se voir attribuer le qualificatif de « données ouvertes ». Ces opérations préalables à toute diffusion constituent un véritable « investissement de forme ». Comme le souligne Laurent Thévenot, « *la formule d'investissement met en balance un coût et la généralité d'une forme qui sert d'instrument d'équivalence et qui est caractérisée par sa stabilité et son extension (domaine de validité)². » D'une situation de complexité initiale caractérisée par la diversité des producteurs et des données, ces opérations sur les données permettent d'établir une forme stabilisée qui assure une équivalence entre des mondes sociaux hétérogènes. Ce processus s'effectue au travers d'un travail simultané de détachement et d'attachement des données³ : par ces médiations, on défait les liens reliant les données à leur environnement initial pour permettre leur association à de nouveaux utilisateurs. Ce mouvement s'inscrit dans une dialectique de la perte et du gain permettant le passage du singulier au général⁴ : perte de certaines propriétés liées à la culture métier pour en gagner*

¹ Dominique Boullier et Maxime Crépel, « Biographie d'une photo numérique et pouvoir des tags. Classer/circuler », *Revue d'anthropologie des connaissances*, 2013, vol. 7, n° 4, p. 785-813.

² L. Thévenot, « Les investissements de forme », art cit, p. 30.

³ Sarah Labelle et Jean-Baptiste Le Corf, « Modalités de diffusion et processus documentaires, conditions du "détachement" des informations publiques. Analyse des discours législatifs et des portails open data territoriaux. », *Les enjeux de l'information et de la communication*, 2012, n° 2012/1, p. 209-222.

⁴ B. Latour, « Le topofil de Boa Vista ou la référence scientifique -montage photo-philosophique », art cit ; B. Latour, « Les "vues" de l'esprit », art cit.

d'autres qui les font gagner en mobilité et en capacité de combinaison. Ces processus de traduction des données permettent le passage d'un univers à un autre.

Toutefois, comme le souligne la formule de Laurent Thévenot, la production des « données ouvertes » est un véritable investissement technique, financier et humain suscitant des réticences de la part d'agents de l'administration. D'une part, parce que ces opérations chronophages s'ajoutent à leur travail quotidien et qu'elles ne sont ni encouragées ni valorisées par leur direction.

Pour l'instant on a rien. Pourquoi on a rien ? C'est un peu lié à un problème d'effectif au sein de la direction de l'eau. On a le nez dans le guidon, et on n'a pas le temps de prendre du recul sur quelles données on pourrait mettre à disposition. Et ensuite, il faut que l'on monte au créneau de notre direction pour leur rappeler l'obligation de diffuser des données publiques. [...] Moi, je suis partisan de mettre à disposition les données, mais pour ce faire, il faut que j'ai les moyens de le faire, c'est-à-dire, définir les métadonnées, classifier nos données. [...] Aujourd'hui, je n'ai pas le temps¹.

D'autre part, ces opérations modifient les modes de faire de l'administration.

Mais, je te dis pour la donnée sur les maisons du Rhône, j'ai halluciné de la résistance des services. On a été voir le service en charge qui nous a fourni un fichier excel, parce que c'était le plus simple pour eux, jusque là, on a pas eu trop de problèmes. Mais ensuite, on a voulu ajouter à ce fichier des informations géographiques, pour pouvoir afficher une carte avec un point pour chaque maison du Rhône et les informations attributaires associées. Et là, ça a mis un temps monstrueux, ça a bloqué de partout, avec chaque acteur qui faisait de la rétention d'information, des services qui ne voulaient même pas faire ce qu'on leur demandait, parce que ça changeait leur mode de faire. Donc, ça a été hyper compliqué. Au lieu de prendre un mois comme je pensais au début, la publication a pris plus de cinq mois. Tout ça pour une nouvelle façon de faire, qui change ce qu'ils font d'habitude².

A l'issue de cette troisième épreuve de diffusibilité, certaines données identifiées et publicisées ne sont pas mises à disposition du fait de contraintes techniques ou organisationnelles. Plus généralement, cette épreuve précise encore le public associé aux données. Certains attributs sont sélectionnés ou modifiés, des formats sont identifiés, des modes d'accès et d'actualisation sont mis en place en fonction d'usages préfigurés des données. Un certain type d'utilisateur est ancré dans l'infrastructure de diffusion au travers des formats, des métadonnées, des modalités d'accès, etc.

¹ Source : Entretien Thibaut, administrateur de données, Grand Lyon (4 décembre 2013)

² Source : Journal de terrain, discussion personnelle avec Léa (21 octobre 2015)

IV. Conclusion – De la « donnée » à la « donnée ouverte »

Ce chapitre a débuté par une interrogation sur ce qu'était une donnée. Au cours des différentes épreuves du processus d'ouverture, les acteurs définissent successivement cette entité. L'épreuve d'identification a permis de catégoriser ce qu'était une « donnée candidate ». La publicisation a fait émerger la « donnée publiable ». Enfin, l'extraction a précisé ce qu'était une « donnée ouverte ». La donnée est modifiée tout au long de la chaîne de diffusion pour finir par se stabiliser comme une entité diffusable. Ce processus est réversible : des données non catégorisées comme telles peuvent le devenir, et, inversement, des données ouvertes peuvent être redéfinies comme non-diffusables.

Mais, à l'issue de ces épreuves, est-ce que la donnée est toujours la même¹ ? A mesure que la donnée circule d'une épreuve à l'autre, elle change non seulement de signification, mais elle change également d'identité. Ce n'est pas différentes interprétations ou catégorisations d'une même donnée, mais une donnée différente par le réseau qu'elle déploie. A chaque épreuve, on pose la question « qu'est-ce qu'une donnée ? », et à chaque fois, la réponse est différente, puisque les médiations qu'on ajoute à la donnée transforment les contours de son réseau.

Plutôt que de considérer les données comme des entités stables et homogènes, j'ai souhaité dans ce chapitre les considérer de manière dynamique, comme des réseaux qu'il convient de déployer pour envisager la singularité de chacune d'entre elles. La donnée comme « réseau » n'a pas de contours fixes, ceux-ci varient avec l'extension et la transformation du réseau, recomposé au prix d'une série d'épreuves. La donnée n'est ainsi pas pré-existante à son ouverture. Au début du processus de diffusibilité, la donnée n'existe pas. Dès lors, il est impossible d'affirmer que la donnée a une essence, c'est-à-dire certaines propriétés desquelles, il serait possible de déterminer, dès le début de ce processus, sa diffusibilité. Loin d'être joué à l'avance, le processus d'ouverture des données est le résultat d'une série

¹ Pour un questionnement similaire sur une entité tout autre, voir Emilie Gomart, « Methadone: Six effects in Search of a Substance », *Social Studies of Science*, 2002, vol. 32, n° 1, p. 93-135.

d'épreuves, au résultat toujours incertain, au cours desquelles les caractéristiques des données, des producteurs, des utilisateurs, sont jugées et redéfinies par les acteurs. Ces différentes *épreuves de diffusibilité* conduisent ainsi à recomposer le réseau des données afin qu'elles soient considérées comme « diffusables ».

Pour comprendre la mise en circulation des données, il faut s'intéresser tout autant aux producteurs, aux dispositifs de production et de stockage de la donnée, aux textes législatifs et réglementaires, aux coûts, aux élus, à la direction de la communauté urbaine de Lyon, aux relations antérieures entre acteurs, aux volontés de pouvoir des uns et des autres, aux formats des données, etc. Face à la multiplicité de ces actants, les épreuves que j'ai décrites dans ce chapitre tentent de forger un consensus et d'aligner les intérêts de chacun d'entre eux. Cela ne peut se faire qu'au prix d'un intense processus de traduction qui voit se redéfinir les identités de l'ensemble de ces actants¹. L'enrôlement des producteurs dans le processus de diffusion s'effectue par la préfiguration d'utilisateurs des données ouvertes. Ces êtres intermédiaires prennent des identités multiples selon les données et les enjeux auxquels on les associe. Les données sont placées dans des situations diverses, aux mains d'acteurs variés, on essaye d'imaginer les avantages, les risques, les inconvénients, les obstacles à tel ou tel usage de la donnée. La « diffusibilité » d'une donnée est l'adéquation entre certaines caractéristiques possédées (ou potentiellement possédées) par une donnée et les usages (réels ou supposés) qui seront faits de la donnée.

Ce travail de traduction participe à un processus conjoint, toujours en tension, de détachement de la donnée de son infrastructure informationnelle dans laquelle elle est insérée et de son attachement à un nouvel environnement. Cependant, à la différence des processus d'innovation étudiés par Goulet et Vinck², la mise à disposition des données ne s'inscrit pas dans une dissociation de l'ensemble des liens entre la donnée et son environnement initial. Ceux-ci sont préservés. En premier lieu car la donnée continue à être produite et utilisée quotidiennement par les services de chacune des institutions. En second lieu, car l'attachement à de nouveaux utilisateurs ne peut se produire si la donnée est entièrement détachée de son environnement initial : la donnée n'est plus mise à jour, certaines

¹ M. Callon, « Éléments pour une sociologie de la traduction. », art cit.

² F. Goulet et D. Vinck, « L'innovation par retrait. », art cit.

informations ne permettent pas son utilisation, etc. Dès lors, tout l'enjeu du processus d'ouverture des données est de parvenir à gérer cette tension entre détachement et attachement. Est-ce que la donnée est suffisamment détachée ? A contrario, est-ce qu'elle n'est pas trop détachée ? La robustesse de l'attachement des données à ces nouveaux publics est l'objet du chapitre suivant.

Encadré 11 : L'open data par défaut facilite-t-il la diffusibilité des données ?

Plusieurs militants de l'*open data* plaident pour la mise en place d'un principe d'*open data* par défaut. L'ouverture deviendrait alors la règle, la fermeture l'exception. En mars 2012, le maire de New-York, Michael Bloomberg, a signé une loi en ce sens : toutes les données de l'administration municipale doivent être diffusées sur le portail *open data* de la ville. Si cette loi a indéniablement favorisé l'ouverture des données, sa mise en œuvre est plus complexe que son postulat originel. En effet, les caractéristiques de diffusibilité des données ne se décrètent pas mais doivent être acquises au cours d'épreuves de diffusibilité. Les données new-yorkaises subissent les mêmes épreuves d'identification, de publicisation et d'extraction qui restreignent leur ouverture.

En premier lieu, elles ne préexistent pas à leur mise à disposition. Les problématiques d'identification des données produites par les différents services de la ville sont similaires à celles du Grand Lyon. Pour reprendre les termes de l'ancien *Chief Data officer*, la ville de New York « *doesn't know what it knows* ». Il n'existe pas d'inventaire référençant l'ensemble des données municipales, celles-ci ne seront que partiellement connues qu'au terme d'un long travail d'identification auprès des services.

En second lieu, la publicisation est également le résultat d'associations de la donnée à de nouveaux utilisateurs. Le responsable de la police de New York est par exemple réticent à mettre à disposition ses données "*which, at best, would serve no purpose and, at worst, would mislead the public. This information is only valuable to those with the training, knowledge*

and experience to understand its context and interpret it correctly. That is the role of the police commander¹.”

Enfin, les processus d'extraction constituent tout autant un ensemble de complexes opérations de détachement. Ils nécessitent des ressources spécifiques qui peuvent ralentir considérablement la diffusion des données municipales.

Comme le révèle le cas de la ville de New York, dont le nombre de données diffusées reste limité², l'impensé du principe d'*open data* par défaut est que les données ne peuvent être ouvertes par défaut, puisqu'elles acquièrent précisément ces caractéristiques de diffusibilité au cours du processus d'ouverture.

¹ Source : James Tuller, *Chief of transportation of the NYPD*, cité dans "Bill to Release Street Safety Data Gains Steam Over NYPD Objections" (28 avril 2010). Disponible en ligne : <http://www.streetsblog.org/2010/04/28/bill-to-release-street-safety-data-gains-steam-over-nypd-objections/> (Consulté le 14 août 2015)

² Environ 1300 jeux de données sont publiés sur la plateforme *open data* au 10 mai 2016, ce qui est considérable en comparaison des portails des municipalités françaises, mais à relativiser par rapport au nombre de données produites et utilisées par l'administration newyorkaise.

Chapitre 4

Réutiliser : Le pluralisme des politiques de réutilisation des données ouvertes

Poursuivons la chaîne des données en nous intéressant à leur « seconde vie » : leur utilisation au sein d'un nouvel environnement. Peu de travaux académiques existent sur les réutilisations des données en *open data*. L'accent est davantage mis sur l'accès aux données plutôt que sur les usages qui en sont faits. Alors que, comme le rappelle Rob Kitchin, les partisans de l'*open data* promeuvent les bénéfices démocratiques de l'ouverture des données¹, l'étude réalisée par Ben Worthy des usages des données budgétaires mises à disposition par les administrations locales britanniques met en avant une utilisation réservée essentiellement aux acteurs disposant de connaissances en finances publiques et ayant une activité de contrôle des activités gouvernementales (médias, activistes, entreprises)². Comme d'autres auteurs l'ont pointé³, ce travail souligne que les citoyens ont un accès inégal au temps, aux compétences et aux ressources nécessaires pour accéder aux données et les analyser.

Dès lors, plutôt qu'une désintermédiation créée par l'accès libre aux données, l'*open data* soulèverait un risque de recomposition du pouvoir au profit de nouveaux intermédiaires et/ou des puissants accentuant les inégalités (« *empower the empowered* »)⁴. Dans cette perspective critique, la chercheuse britannique Jo Bates souligne que le mouvement de l'*open data* renforce les intérêts des entreprises et de la finance⁵. Analysant le cas du Royaume-Uni, elle affirme que ce mouvement en faveur de la diffusion des données s'est accéléré lorsque des entreprises s'en sont faites les promotrices auprès du gouvernement, lequel y voyait un intérêt pour poursuivre sa politique d'austérité et de marchandisation des services publics. Les acteurs étatiques se sont appropriés « *the initiative on behalf of dominant capitalist interests*

¹ Rob Kitchin, *Four critiques of open data initiatives*, <http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/>, 27 novembre 2013.

² Ben Worthy, « The Impact of Open Data in the UK: Complex, Unpredictable and Political », *Public Administration*, 2015, vol. 93, n° 3, p. 788-805.

³ Kevin P. Donovan, « Seeing Like a Slum: Towards Open, Deliberative Development », *Georgetown Journal of Informational Affairs*, 2012, vol. 13, n° 1, p. 97-104 ; Tom Slee, *Seeing like a geek.*, <http://crookedtimber.org/2012/06/25/seeing-like-a-geek/>, 25 juin 2012, (consulté le 6 octobre 2015).

⁴ Michael Gurstein, « Open data: Empowering the empowered or effective data use for everyone? », *First Monday*, 2011, vol. 16, n° 2.

⁵ Jo Bates, « The Strategic Importance of Information Policy for the Contemporary Neoliberal State: The Case of Open Government Data in the United Kingdom. », *Government Information Quarterly*, 2014, vol. 31, n° 3, p. 388-395.

*under the guise of a 'Transparency Agenda'*¹. » Selon Jo Bates, l'*open data* s'inscrit dans la continuité des réformes du secteur public et du *new public management*². En encourageant la mise en concurrence des services publics, l'ouverture des données publiques conduirait ainsi à une recomposition du pouvoir au profit d'acteurs privés³.

Ces travaux, qui s'inscrivent dans une perspective critique, analysent les recompositions du gouvernement en se centrant sur les nouvelles configurations d'acteurs qui émergent avec l'ouverture des données. En l'absence d'études empiriques, ils reposent sur le postulat implicite, que si des données sont mises à disposition, des acteurs vont s'en emparer, les utiliser et en retirer de la valeur. La logique associant *open data* et développement économique est basée sur la même assertion : partager les données va permettre à d'autres d'en retirer de la valeur (« unlock their value »). Or, ce modèle selon lequel une offre de données publiques rencontrerait automatiquement une demande de données ne résiste pas à l'analyse. Loin des promesses initiales de développement économique et de renouvellement démocratique, à l'heure d'un premier bilan des démarches *open data*, les acteurs du secteur portaient un regard désappointé face à la faiblesse (relative) des réutilisations des données ouvertes. Une observatrice attentive de l'*open data* en France témoignait ainsi en janvier 2013 d'un « baby blues de l'*open data* » : « *les organisations traversent une phase de déprime et de doute, le data blues* » ; « *la multitude de difficultés techniques, juridiques, culturelles, et organisationnelles a laissé un goût amer aux réutilisateurs de données.* »⁴. Ainsi, ce n'est pas parce qu'une donnée est mise à disposition qu'elle sera (ou pourra) être réutilisée. Quels sont les défis auxquels sont confrontés les réutilisateurs de données ? Comment font-ils pour les surmonter ? Ces questions soulèvent plus largement des enjeux de communication et de

¹ Jo Bates, « “This is what modern deregulation looks like”: Co-optation and contestation in the shaping of the UK's Open Government Data Initiative. », *The Journal of Community Informatics*, 2012, vol. 8, n° 2.

² Justin Longo, « Open Data: Digital-Era Governance Thoroughbred or New Public Management Trojan Horse? », *Public Policy and Governance Review*, 2011, vol. 2, n° 2, p. 38-51 ; Amanda Clarke et Helen Margetts, « Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform », *Policy & Internet*, 2014, vol. 6, n° 4, p. 393-417.

³ Justin Keen et al., « Big data + politics = open data: The case of health care data in England », *Policy & Internet*, 2013, vol. 5, n° 2, p. 228-243 ; J. Bates, « The Domestication of Open Government Data Advocacy in the United Kingdom: A Neo-Gramscian Analysis », art cit.

⁴ Source : <https://libertic.wordpress.com/2013/09/24/vers-la-fin-du-baby-blues-de-lopen-data/> (Consulté le 14 mai 2016)

coordination entre des mondes sociaux hétérogènes¹. Comment les données produites et utilisées pour répondre aux besoins de l'administration peuvent-elles être utilisées au sein d'autres univers sociaux ?

Pour répondre à ces interrogations, plutôt que de se focaliser uniquement sur les réutilisateurs, il est nécessaire de faire entrer d'autres entités dans l'analyse, dont en premier lieu les données elles-mêmes. Loin d'être neutres, celles-ci ne se laissent pas aisément réutiliser. De multiples opérations de nettoyage, de croisement, de standardisation, d'articulation, etc. sont nécessaires pour permettre leur attachement à un nouvel environnement. Pour analyser ce que changent les réutilisations au gouvernement de la ville, il faut s'intéresser au statut de chacune de ces médiations qui redéfinissent les liens entre les actants du réseau sociotechnique. Certaines entités sont associées, d'autres dissociées, transformant d'autant les données. Tout l'enjeu est alors de réussir à aligner ces médiations pour « faire tenir » le réseau de ces données ouvertes et assurer leur solidité.

Suivre les données offre également l'avantage de ne pas se restreindre à une analyse « externaliste » de l'*open data*, mais d'être attentif aux évolutions des représentations de l'espace dont sont porteuses les données. En effet, associer de nouvelles médiations à la donnée pour permettre sa réutilisation dans un nouvel environnement n'est pas neutre : en quoi l'utilisation de la donnée au sein d'un autre monde social transforme les représentations de l'espace urbain ? Comment ces représentations sont-elles négociées et stabilisées entre mondes sociaux ? Quelle est *in fine* leur influence sur les modalités de gouvernement de la ville ?

Au travers de l'étude de plusieurs réutilisations de données mises à dispositions par le Grand Lyon, ce chapitre révèle différentes manières de composer avec les données ouvertes. Les enjeux socio-politiques variés qui émergent de cette pluralité de modalités de réutilisation

¹ La notion de monde social est empruntée au courant de l'interactionnisme symbolique. Elle désigne des « groupes en activités, n'ayant nécessairement ni frontières claires ni organisation solide. », qui se constituent dans la trame des interactions sociales issues de leur activité primaire et de la définition de la réalité pertinente. (Anselm Strauss, *La trame de la négociation: Sociologie qualitative et interactionnisme*, Paris, L'Harmattan, 1992, p. 271.)

Ici, je n'étudie pas tant la formation de mondes sociaux que les modalités de coordination entre ceux-ci par l'intermédiaire des données ouvertes.

des données, soulignent le pluralisme des politiques de réutilisation. Elles sont présentées dans trois catégories selon les opérations qu'elles font subir aux données : la consolidation, l'homogénéisation et l'articulation. Ces trois familles de réutilisation, auxquelles sont associés de nouveaux agencements de la chaîne sociotechnique de la donnée, ne sont pas exclusives l'une de l'autre, mais coexistent.

Encadré 12 : Comment identifier les réutilisateurs des données ?

Le libre accès aux données rend complexe la compréhension des usages effectifs des données et l'identification des réutilisateurs. Le Grand Lyon dispose de peu de statistiques sur la fréquentation et l'usage de la plateforme de diffusion de données. Depuis l'ouverture de cette dernière¹, environ 60 000 visiteurs se sont rendus sur la plateforme, soit une moyenne de 70 visiteurs uniques par jour. La faiblesse relative² de ces chiffres de fréquentation s'explique partiellement par le fait que, d'une part, la plateforme est destinée à un public restreint de professionnels disposant des capacités pour utiliser les données, et d'autre part, de nombreux utilisateurs ne visitent qu'une fois le site et automatisent ensuite le téléchargement des données. L'analyse plus précise de ces chiffres de fréquentation révèle que les visites sont près de trois fois plus importantes les jours ouvrés que les weekends³. Cela souligne l'utilisation essentiellement professionnelle de la plateforme de diffusion de données. Le profil des réutilisateurs qui émerge de l'analyse des emails reçus par l'équipe en charge de la plateforme *open data* (voir annexe) confirme cette utilisation majoritaire dans un cadre professionnel⁴. Les citoyens ou les organisations de la société civile semblent peu s'emparer des données mises à disposition par le Grand Lyon.

Le Grand Lyon ne dispose pas de statistiques précises sur le nombre de téléchargement ou d'accès via les API à chacune des données. La seule information connue est celle du

¹ Les chiffres présentés couvrent la période du 31 mai 2013 au 6 octobre 2015.

² A titre de comparaison, le site *grandlyon.com* reçoit en moyenne 4000 visiteurs uniques par jour.

³ Du lundi au vendredi, la plateforme enregistre en moyenne 87 visites uniques quotidiennes, contre 29 les samedi et dimanche.

⁴ Ce profil majoritairement professionnel du réutilisateur est identique pour les 11 données mises à disposition après authentification des utilisateurs. Parmi les 51 signataires d'une convention d'utilisation de ces données, 45% sont des entreprises, 40% des personnes privées et 15% des universités et des institutions publiques.

nombre de consultations de la fiche de données sur la plateforme. Les variations sont importantes entre les données. La donnée de l'état du trafic en temps réel est la plus consultée (12065 visites au 12 octobre 2015), tandis que la maquette 3D de la commune de Saint-Germain-au-Mont-D'or n'a reçu que 15 visites.

A partir de ces quelques chiffres, pour mener l'analyse de ce chapitre, j'ai sélectionné une quinzaine de cas représentatifs de la variété des acteurs et de la diversité des usages identifiés des données ouvertes du Grand Lyon¹. En premier lieu, j'ai souhaité étudier des réutilisations de données géographiques et de données mobilité. En second lieu, j'ai varié les secteurs dans lesquels interviennent ces réutilisateurs : des acteurs publics ont été interrogés², dont notamment les usages internes à la communauté urbaine de Lyon et à la ville de New York³ ; une organisation de la société civile (l'association OpenStreetMap) ; et des entreprises intervenant dans le secteur de la mobilité⁴, du traitement de données géographiques⁵ et de la modélisation urbaine⁶. Si les tailles de ces entreprises sont variées, de la start-up au groupe international en passant par la PME, il faut préciser que les données constituent le cœur de l'activité économique de la quasi-totalité d'entre elles⁷.

I. Consolidation

L'utilisation secondaire d'une donnée soulève des enjeux de coordination entre mondes sociaux hétérogènes. Marquée de l'empreinte institutionnelle de son producteur, la donnée ouverte ne se laisse pas aisément utiliser au sein d'autres univers organisationnels ou

¹ La sélection a été effectuée à partir de contacts établis par l'équipe du projet *open data* au travers de mails de réutilisateurs reçus et des utilisateurs ayant signé des licences avec authentification.

² Lyon, Villeurbanne, Vénissieux, Vaulx-en-Velin, Chassieu, Caluire et Cuire, Tassin-la-Demi-Lune, ainsi que l'Agence d'urbanisme de Lyon.

³ Les projets Optimod, Transform et l'analyse du MODA.

⁴ Canal TP, Cityway, Renault Trucks, Géovélo, IBM, Google.

⁵ Oslandia, Nokia Here.

⁶ ForCity.

⁷ Seule l'entreprise Renault Trucks ne base pas (aujourd'hui ?) son activité économique sur des produits ou des services de données.

professionnels. La dissociation des liens entre production et usage met en péril la solidité de la donnée. Puisqu'on ne peut s'assurer que les modalités de production de la donnée répondent aux besoins des réutilisateurs, les données sont menacées de déliquescence¹.

Face à ces risques de déliquescence, tout l'enjeu est de renforcer la solidité de la donnée, c'est-à-dire sa capacité à résister à des circonstances et des pratiques inattendues. A la suite de Eugène Dupréel, Emmanuel Didier souligne que la solidité d'une donnée ne repose pas sur sa stabilité, mais sur l'allongement de la chaîne de cette donnée par l'ajout d'éléments permettant le passage d'un point A à un point B². Ces opérations de consolidation renforcent la consistance de la donnée ouverte et autorisent son usage au sein d'un nouvel environnement.

Revenant sur des cas de déliquescence de données ouvertes puis sur ces processus de consolidation, la première partie de ce chapitre présente l'évolution de la donnée ouverte vers le statut d'objet-frontière. Le concept d'objet-frontière a été proposée par Susan L. Star et James R. Griesemer pour désigner les objets qui sont « *both adaptable to different viewpoints and robust enough to maintain identity across them* »³. La donnée ouverte comme objet-frontière doit ainsi être suffisamment plastique pour s'adapter aux pratiques hétérogènes des réutilisateurs, tout en restant suffisamment robuste pour maintenir une représentation commune de l'espace urbain entre les acteurs. Pour cela, elle doit acquérir deux propriétés principales : une flexibilité interprétative et une infrastructure partagée (des conventions, des normes ou des formats attachés à une communauté de pratiques)⁴. Ces deux caractéristiques ne sont pas en soi attachées aux données, mais elles dépendent des situations dans lesquelles

¹ Emmanuel Didier définit la déliquescence comme la perte de prévisibilité des liens qui unissent les éléments qui composent un agrégat.

Emmanuel Didier, « La consistance du futur. La prévision statistique aux États-Unis et la Grande Dépression », *Raisons politiques*, 2012, vol. 4, n° 48, p. 65-83.

² E. Didier, *En quoi consiste l'Amérique ? Les statistiques, le New Deal et la démocratie*, op. cit.

³ L'objet-frontière peut prendre différentes formes. Star et Griesemer en présente quatre : les répertoires, les idéaux-types, les frontières conventionnelles et les formats standardisés.

S.L. Star et J.R. Griesemer, « Institutional Ecology, "Translations" and Boundary Objects », art cit.

⁴ L'articulation entre ces deux propriétés de l'objet-frontière est particulièrement mis en avant dans le dossier de la revue d'anthropologie des connaissances consacré au concept d'objet-frontière. Pascal Trompette et Dominique Vinck, « Retour sur la notion d'objet-frontière », *Revue d'anthropologie des connaissances*, 2009, vol. 3, n° 1, p. 5-27.

sont placées les données ouvertes, des acteurs qui les mobilisent et des pratiques qui leur sont associées.

A. Quelques situations de déliquescence des données ouvertes

Les données mises à disposition sur l'infrastructure de diffusion sont celles utilisées quotidiennement par les institutions productrices. Ces données restent marquées de leur empreinte institutionnelle. Les usages initiaux sont inscrits dans le format et la structure des données limitant le champ d'utilisation possible de ces données. Lorsqu'on souhaite utiliser ces données pour des usages alternatifs, la solidité des données est mise à l'épreuve faisant courir le risque de leur déliquescence.

1) Le manque d'informations contraint l'usage alternatif de la donnée

Le premier critère de déliquescence de la donnée est l'absence, dans le jeu de données initial, des informations nécessaires pour une utilisation alternative de la donnée. Par exemple, la base de données « Voies et adresses », constituée de la trame viaire (le filaire de voies), des numéros de voiries (adresses) et de la toponymie (nom de la voie), constitue l'outil principal de localisation¹ du système d'information géographique. Cependant, son usage dans une optique autre que la localisation, comme le calcul de navigation² par exemple, est impossible du fait du manque d'informations de la base de données. « *Concrètement, si tu mets ça dans un outil type GPS, ce n'est pas adapté, ce n'est pas ce qu'on voulait faire à la base. Il manque toutes les infos de circulation en fait : j'ai pas le droit de tourner à droite, ou le haut gabarit, il y a un pont, etc. Enfin, tout un tas d'informations...* »³. En l'absence de graphe de circulation, cette base de localisation ne permet ainsi pas la navigation.

Ces données sont inexploitable pour pouvoir faire du parcours de graphe, c'est-à-dire pour pouvoir aller d'un point à un autre, parce qu'il n'y a pas les objets nœuds qui permettent de dire "de là à là, on peut y aller, de là à là, on ne peut pas y aller". On pourrait croire que dans les données CRITER dont ils se servent pour réguler le trafic, il y a l'information qui permet de savoir

¹ La localisation consiste à déterminer la position géographique d'un objet (ses coordonnées : latitude/longitude).

² La navigation consiste à calculer, à partir d'une position déterminée, l'itinéraire à suivre pour rejoindre un autre point de coordonnées connues. La navigation est ainsi impossible sans localisation préalable.

³ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

comment on va d'un point a à un point b, et en fait, elle n'y est pas. [...]Comme ils ne font pas l'effort, de rendre leurs données navigables, le résultat c'est que l'on ne peut pas les utiliser.¹

Un autre exemple de ces difficultés de réutilisations, interne à l'administration, s'est manifesté lors de la refonte des sites du Grand Lyon. Il s'est avéré impossible d'utiliser les données en *open data*, car elles ne contenaient pas certaines informations indispensables pour un usage d'information à destination du public.

On a voulu utiliser les données de SmartData pour alimenter les sites intranet et Grandlyon.com lors de la refonte pour la métropole, mais on n'a pas pu, car il y a des données, comme les déchetteries, qui ne sont pas du tout adaptées pour l'utilisateur final, on a l'emplacement des déchetteries, mais pas leur numéro de téléphone ou leurs horaires d'ouverture. On a donc du créer nos propres bases de données, intégrées à chacun des sites².

Cet exemple démontre l'ancrage de la donnée dans un usage particulier, son attachement à un contexte particulier, duquel il n'est possible de s'extraire. La donnée est le reflet de la culture-métier du producteur. L'attachement de la donnée à son cadre d'action initial est tellement fort que la donnée se révèle inadéquate pour un autre usage.

2) L'échelle est limitée au périmètre institutionnel du producteur

Le deuxième facteur de déliquescence de la donnée est l'échelle de la donnée, qui correspond au périmètre de compétence de l'institution. Or, en fonction du découpage institutionnel et d'arrangements locaux, les données peuvent être très incomplètes sur un objet particulier. Par exemple, le Grand Lyon dispose d'un jeu de données des bornes fontaines de son territoire. Mais, celui-ci ne contient pas, entre autres, les fontaines disposées dans le parc de la Tête d'or, car le réseau d'approvisionnement en eau des parcs et jardins n'appartient pas à la domanialité du Grand Lyon. « *Dès que vous êtes dans un domaine privatif, ou un aménagement qui est géré par une copropriété, s'il y a une borne fontaine, nous on n'y aura pas accès, parce que c'est du réseau privatif et pas du réseau communautaire*³ ». Autre exemple, un utilisateur s'interroge : pourquoi n'y a-t-il que deux toilettes publiques sur la ville de Lyon listées dans le jeu de données présent sur la plateforme ? Parce que le Grand

¹ Source : Entretien avec Jacques, chef de projet, IBM (27 janvier 2015)

² Source : Journal de terrain, discussion avec Grace (19 janvier 2015)

³ Source : Entretien avec Thibaut, administrateur de données, Grand Lyon (4 décembre 2013)

Lyon ne recense que les toilettes qu'il doit nettoyer. Les toilettes automatiques situées sur le territoire de la ville de Lyon n'entrent pas dans ce cadre, et ne sont donc pas recensées dans le jeu de données disponible en open data. Ainsi, l'échelle du jeu de données en *open data* correspond au périmètre de compétence de l'institution, délimité par définition par des frontières administratives.

Un point important concerne l'emprise géographique de la donnée. Par exemple, sur les aires de livraison, on a l'information uniquement sur la ville de Lyon, pas sur tout le territoire du Grand Lyon. On a zéro informations sur Villeurbanne, alors que pour nos livreurs, il faudrait que l'on en ait partout. Je n'ai pas insisté là-dessus auprès du Grand Lyon, parce que j'ai vu à quel point c'était compliqué de récupérer des données auprès des villes et de les agréger, donc aujourd'hui, ils ne publient que ce qu'ils ont¹.

La limitation de l'emprise géographique des données au périmètre de compétences de chaque institution empêche l'usage des données sur un espace plus large que ces territoires.

3) La structuration hétérogène des données

Le troisième facteur de déliquescence est lié à la structuration de la donnée. En l'absence de normes nationales ou internationales - ce qui est le cas pour la quasi totalité des données - la structure interne et la granularité de la donnée sont variables d'une institution productrice à une autre. Le modèle de données du plan d'occupation des sols du Grand Lyon n'est, par exemple, pas le même que celui de la communauté urbaine de Bordeaux ou de la métropole de Nice.

Chacun fait à sa sauce. Une bête occupation des sols, elle n'est pas transposable, enfin, elle est rarement transposable d'une région sur l'autre, et ça c'est que la France. [...] La donnée, elle ne se plugue jamais directement dans nos modélisations. Ça arrive une fois sur un million ! Tout le travail de la cellule géomatique au sein de la société, c'est de faire en sorte que ça puisse rentrer dans des modèles de manière normée. La question de l'hétérogénéité, c'est presque notre raison de vivre ! (rires)²

Les données produites au sein de chaque organisation sont autant de conventions d'équivalence institutionnellement situées qu'il faut reconstruire si l'on veut les utiliser à une échelle différente. L'absence d'homogénéisation des données oblige les réutilisateurs à effectuer un travail d'intégration des données spécifique pour chacune des institutions

¹ Source : Entretien téléphonique avec Jean-Paul, chef de projet, Renault Trucks (13 février 2015)

² Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

productrices. Ce coût supplémentaire restreint le développement de services à partir de ces données dont la rentabilité ne serait atteinte, a minima, qu'à l'échelle nationale. L'impossibilité de réaliser ces économies d'échelle limite l'utilisation des données ouvertes à des expérimentations, sans développement commercial ultérieur, comme témoignent ces deux réutilisateurs.

N'empêche que moi je peux faire beaucoup de travail en expérimental au niveau du Grand Lyon, mais le Grand Lyon peut faire tout ce que je veux, jamais je n'aurai un produit qui exploite ses données tant que ce ne soit pas disponible au moins sur toute la France. Il faut que ce soit harmonisé au niveau de toutes les grandes villes pour qu'au moins je puisse exploiter les données de Lyon telles qu'elles sont présentées par Lyon. [...] Aujourd'hui, si je voulais déployer le système *SmartDeliveries* dans une autre ville de Lyon, j'aurai tout un travail spécifique à réaliser pour chaque ville en question. Ce n'est pas viable¹.

Après, le produit n'a pas été développé et commercialisé au-delà de l'expérimentation pour des raisons financières. On avait réfléchi au business case, et l'on voit assez vite que ce n'est pas viable pour nous de développer uniquement le service pour le Grand Lyon. On s'adresse à des livreurs qui peuvent être partout en France, pas seulement sur l'aire lyonnaise. [...] Pour être rentable, il faudrait que l'on déploie le service à l'échelle nationale pour le proposer dans plusieurs villes. Mais, là le problème, c'est que, si le protocole pour récupérer les données est similaire d'une ville à l'autre, les structures des données sont différentes. Elles ne sont pas normalisées, elles sont spécifiques à chaque ville. J'avais regardé un peu la donnée aire de livraison proposée par Bordeaux, et elle n'était pas structurée de la même façon. Donc, ça demanderait chez nous un travail d'intégration spécifique pour chaque ville, qui a un coût non négligeable. Aujourd'hui, ce n'est pas rentable².

*

* *

Informations manquantes, échelle limitée, structuration hétérogène, ces trois caractéristiques sont autant de facteurs de déliquescence des données ouvertes. La dissociation des activités de production et d'usage rend imprévisible la chaîne de la donnée. Comme le remarque Christine Borgman dans son étude des échanges de données scientifiques, plus l'usage secondaire s'éloigne de l'usage initial, que ce soit en termes de temporalité ou de discipline, plus les risques de déliquescence de la donnée sont élevés. « *The farther afield that the reuse is from the point of origin, whether in terms of time, theory,*

¹ Source : Entretien avec Jacques, chef de projet, IBM (27 janvier 2015)

² Source : Entretien téléphonique avec Jean-Paul, chef de projet, Renault Trucks (13 février 2015)

discipline, or other measure of distance, the more difficult it may be to interpret a dataset or assess its value for reuse »¹.

Ces risques de décomposition des données mettent en lumière leur héritage institutionnel. Produites pour répondre à des besoins spécifiques, au sein d'un environnement particulier, les données ouvertes sont difficilement réutilisables dans un cadre d'action alternatif. Les données sont marquées d'une empreinte institutionnelle qui contraint leurs usages ultérieurs. Cela explique la relative faiblesse des réutilisations économiques des open data des collectivités territoriales.

Il existe toutefois de multiples exemples de réutilisation des données ouvertes. Comment les données parviennent à passer d'un monde social à un autre tout en préservant leur solidité ? Pour devenir des objets-frontières, assurant la coordination entre des mondes sociaux hétérogènes, les données ouvertes doivent être dotées d'une flexibilité interprétative et d'une infrastructure partagée entre ces groupes sociaux. Ces caractéristiques ne sont pas intrinsèques aux données mais dépendent des situations dans lesquelles celles-ci sont mobilisées.

B. Opérations de consolidation et estimation de la « flexibilité interprétative » des données ouvertes

Préalablement à tout usage secondaire des données ouvertes, les réutilisateurs réalisent deux opérations de « sourcing » et de « nettoyage ». La première consiste à identifier, comprendre et estimer la fiabilité de la donnée pour l'usage que l'on souhaite en faire. La seconde se compose de l'ensemble des opérations de préparation de la donnée afin de l'intégrer au sein d'un nouvel environnement informationnel. Ces deux opérations renforcent la consolidation des données ouvertes en testant leur flexibilité interprétative, c'est-à-dire en s'assurant qu'elles peuvent être utilisées dans un cadre d'action alternatif sans perdre leur signification initiale.

¹ C.L. Borgman, *Big Data, Little Data, No Data*, op. cit., p. 13.

1) Le « sourcing » : identifier, comprendre et estimer la fiabilité de la donnée

La première étape nécessaire à la réutilisation des données est le « sourcing ». Par l'identification, la compréhension et l'estimation de la fiabilité de la donnée, cette opération permet au réutilisateur d'acquérir une familiarité avec la donnée afin d'estimer son usage éventuel au sein d'un nouvel univers social.

Le sourcing passe en premier lieu par une identification de la donnée pertinente sur la plateforme *open data*. En mettant fin aux « pratiques privées¹ » d'échange de données², cette infrastructure a facilité l'accès aux données. Toutefois, l'existence de fichiers à la dénomination analogue rend parfois compliquée l'identification des données pertinentes. Un interlocuteur de l'entreprise de cartographie numérique Here indique son incompréhension face à la multiplicité des fichiers voies et adresses du Grand Lyon.

J'ai quelques soucis sur l'open data, la multiplicité des informations qui sont parfois redondantes. Par exemple, sur le filaire routier, j'ai l'impression qu'il y a plusieurs jeux de données. Et je ne sais jamais lequel je dois prendre ! Par exemple, ce jeu : « Nom de voie du Grand Lyon (Voies et adresses) », on pourrait se dire ce sont les adresses et la voirie, mais si on regarde en dessous, il y en a un autre qui s'appelle « numéros de voirie du Grand Lyon », alors que là, c'est « nom de voies du Grand Lyon ». Et souvent je me suis demandé, mais est-ce que ça change quelque chose ? [...] Donc, je me suis marqué sur un petit post-it le jeu de données que je dois prendre, parce que là ce n'est pas clair.³

Pour identifier le jeu de données, il est souvent nécessaire « d'entrer » dans le jeu de données afin de connaître les informations qu'il contient. Ceci est particulièrement important, dans le cas de jeu de données complexes, tel que celui du PLU, où les informations relatives au zonage sont réparties entre plusieurs données.

¹ L'expression « pratiques privées » est empruntée à Christine Borgman. Elle caractérise ici la situation préalable où chaque producteur était responsable de la diffusion de ses données, ce qui contribuait à une diversité des modalités de diffusion (formats, licences, etc.).

Ibid., p. 227.

² Ces contraintes d'accès aux données pouvaient conduire à des situations ubuesques. Un interlocuteur de l'agence d'urbanisme nous raconte ainsi qu'une personne au sein du service SIG d'une commune du Grand Lyon était chargée de redessiner le PLU à partir de plans à chaque nouvelle version du PLU opposable, alors que les données actualisées sous format SIG étaient disponibles au sein de l'agence urbanisme. Il refaisait ainsi un travail déjà effectué par ailleurs au risque d'erreurs et de divergences entre les différentes versions.

³ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

Il faut prendre le temps de se pencher dessus pour comprendre. Parce que quand on regarde les zonages, il y a une partie de l'information du zonage est dans « contour de zonage », l'autre est dans l'objet représentant l'espace privé à l'intérieur du zonage. C'est là qu'on va récupérer la hauteur d'îlot, le niveau de la bande de caisse, quand il y a des caisses qui sont définies. Pour comprendre entre les deux et compiler, l'ensemble d'un zonage et les informations sur ce que j'ai le droit en terme de hauteur, en terme d'alignement, en terme de caisse, et du coup, sur quel code de règlement je renvoie, c'est pas dans un seul objet, c'est dans plusieurs objets, et donc il faut le savoir quoi.¹

Il est ainsi nécessaire de se familiariser avec les jeux de données. Ce travail de compréhension est indispensable pour s'approprier la donnée et estimer si elle correspond au besoin souhaité. Les métadonnées renseignées par les producteurs prennent ici tout leur sens. Elles permettent d'identifier ce que contient la donnée, comment elle a été produite et les usages que l'on peut en faire.

Néanmoins, à l'instar de la réutilisation des données environnementales étudiées par Ann Zimmerman², les informations contenues dans les métadonnées ne sont pas toujours suffisantes. « *La métadonnée qui va avec [la donnée] va permettre d'avoir une première évaluation de sa pertinence et de pouvoir soutenir une évaluation de qualité, y'a de la métadonnée, y'en a un peu, mais parfois on a besoin de plus* »³. Le Grand Lyon reçoit de nombreux emails de personnes demandant des informations complémentaires sur les données. Ces difficultés de compréhension peuvent relever de la nomenclature utilisée (« *quelles sont les significations des valeurs suivantes pour les états du trafic : "G", "*", "O", "R", "V" ?* »⁴) ou des informations contenues dans le jeu de données :

Sur le portail de données du Grand Lyon, je trouve cet ensemble de données sur les fontaines publiques, mais pourquoi y a-t-il si peu d'informations permettant de mettre du contexte sur ces données ? La seule information de contexte est la fréquence de la mise à jour, mensuelle... Qui laisse penser que les données qui composent cet ensemble est incomplet (m'étonnerait que le Grand Lyon ou ses partenaires (Veolia) construisent de nouvelles fontaines toutes les semaines). Quid des données manquantes ? Il semble n'y avoir aucune fontaine sur Villeurbanne, est-ce parce que les données de Villeurbanne sont absentes ? Pas de Fontaine au parc de la tête d'or, ni sur le plateau de la Croix Rousse ? Est-ce parce qu'il n'y en a pas ou parce que les données ne sont pas à

¹ Source : Entretien avec Olivier, responsable SIG, agence d'urbanisme de Lyon (25 février 2015)

² Ann Zimmerman, « Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse », *International Journal on Digital Libraries*, 2007, vol. 7, n° 1, p. 5-16.

³ Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

⁴ Source : Email reçu par l'équipe en charge de la plateforme d'ouverture de données du Grand Lyon (10 février 2015)

jour ? Ce problème est dénoncé depuis longtemps dans la publication des jeux de données... Mais il les rend assez inutilisables et "inu-réutilisables".... On peut en savoir plus ?¹

Les réutilisateurs doivent ainsi étudier spécifiquement chacune des données afin de déterminer si les informations de la donnée, leurs précisions, leurs fiabilités permettront de répondre à l'usage qu'ils souhaitent en faire.

On rencontre toujours la problématique du sourcing de la donnée : qu'est ce qu'il y a dedans ? Comment elle a été faite ? Mais, après, c'est tellement spécifique à chaque couche de données que la question se pose à chaque fois. La seule ligne directrice que l'on a, c'est : pourquoi je récupère cette donnée, quel est l'usage que je vais en faire et pour cet usage quel est le degré de qualité et quelle est la normalisation dont je vais avoir besoin. Finalement, à partir de là, une donnée va être bonne, une donnée va être rejetée. [...] La problématique principale, c'est de sourcer : la question de la validation, l'évaluation de la précision et de la validité est un truc hyper important.

Enquêteur : Cette fiabilité est liée à l'usage qui est fait de la donnée ?

Elle est totalement liée à l'usage de la donnée, c'est-à-dire qu'en fonction de l'usage que l'on a de la donnée, on n'aura pas du tout la même approche et la même exigence de qualité².

Cette opération de « sourcing » est spécifique à chaque usage de la donnée. Une même donnée peut ainsi être jugée utilisable dans une situation particulière mais ne pas convenir pour un autre usage. Un réutilisateur de données d'un réseau de transport de cars interurbains a eu la surprise découvrir que la localisation de nombreux arrêts était peu précise. Ce défaut de fiabilité ne représente pas un problème important dans le cadre d'un service d'information, mais il est crucial lorsque l'on souhaite mettre en place un calculateur d'itinéraire multimodal.

Il y avait un problème important de positionnement des arrêts. C'était pas important pour eux pour faire de l'information monomodale, mais quand vous faites du calcul d'itinéraire multimodal, c'est indispensable de connaître où sont précisément les arrêts. Il y a une différence importante entre faire de l'information aux voyageurs et faire du calcul d'itinéraire. Dans le premier cas, vous avez un fichier excel avec l'ensemble des données d'horaire théorique, ça vous suffit, mais pour le second cas, il faut que la donnée soit numérisée proprement, le niveau d'exigence n'est pas le même afin de pouvoir d'une part faire du calcul à partir de celle-ci, et d'autre part, de combiner la donnée avec une autre pour faire du multimodal³.

Comme en témoigne cet ingénieur de l'entreprise Cityway, la fiabilité d'une donnée ne peut être estimée qu'en fonction de l'usage qui en sera fait. Les réutilisateurs utilisent plusieurs techniques pour juger du degré de qualité de la donnée. Ils la comparent notamment

¹ Source : Email reçu par l'équipe en charge de la plateforme d'ouverture de données du Grand Lyon (4 juin 2014)

² Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

³ Source : Entretien téléphonique avec Alexis, ingénieur, Cityway (9 juillet 2014)

à d'autres sources d'informations : bases de données similaires, vérification sur Google Street View ou, lorsque cela est possible par un repérage sur le terrain. Par exemple, ces utilisateurs des données de transport en commun identifient des erreurs ou des anomalies dans les jeux de données en les comparant avec les informations fournies sur le site de l'opérateur de transport TCL.

Nous utilisons les données "TCL horaires théoriques" au sein de notre application, et je pense avoir trouvé des erreurs dans le GTFS. Aujourd'hui (lundi 6 avril, le lundi de Pâques) ainsi que le 1er Mai, il n'y a aucun "trip" d'actif. Dans l'application officielle de la TCL, je vois que même si c'est un jour férié, il y a quand même des véhicules qui circulent sur certaines lignes. De même il n'y a que 15 "trips" actifs le 8 mai, ce qui me paraît vraiment peu...¹

Je trouve une incohérence entre les données fournies sur <http://data.grandlyon.com/> et le site tcl.fr.

Voici ma démarche pour illustrer le problème :

- Aller sur <http://www.tcl.fr/>

- Rubrique horaire

- Sélection T3

- Direction Part Dieu

- Arret Gare de Villeurbanne

=> Les deux premiers passages sont à 04h40 et 04h43.

- Aller sur <http://data.grandlyon.com/>

- Télécharger les horaires théoriques du réseau TCL sous format GTFS.

- Ouvrir le fichier `stop_times.txt` de l'archive `GTFS_TCL.zip`

A partir de là, je ne trouve aucune trace de l'horaire 04:43. En revanche l'horaire 04:40 existe bien "n" fois (ex: T3_36_4_070AM_071002,04:40:00,04:40:00,35665,8).

Pouvez-vous m'aider à comprendre ?²

Un autre utilisateur estime la fiabilité de la donnée en la comparant avec sa connaissance de la ville de Lyon ou en utilisant les services de visualisation proposés par Google.

Sur Lyon, on peut limite aller faire du repérage terrain quand on a un doute de positionnement d'une école ; on a un Google Earth qui va éventuellement permettre d'aller repérer le terrain de jeu, devant le bâtiment et d'évaluer les différences par rapport à la « réalité »³.

Le sourcing est le processus de compréhension d'une donnée et d'évaluation de sa qualité. A l'instar des chercheurs en sismologie confrontés à de nouvelles données, étudiés

¹ Source : Email reçu par l'équipe en charge de la plateforme d'ouverture de données du Grand Lyon (6 avril 2015)

² Source : Email reçu par l'équipe en charge de la plateforme d'ouverture de données du Grand Lyon (21 juillet 2015)

³ Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

par Faniel et Jacobsen, le sourcing répond à trois interrogations : « 1) *are the data relevant*, 2) *can the data be understood*, and 3) *are the data trustworthy* »¹. Cette dernière question de la confiance est primordiale pour assurer la réutilisabilité d'une donnée. Si le réutilisateur ne peut s'y fier, il préférera s'en passer plutôt que de baser ses analyses sur une donnée erronée. La réputation du producteur, la précision et la qualité de la documentation de la donnée assurent cette relation de confiance. Un réutilisateur donnera du crédit à une donnée en fonction des informations qu'il aura pu réunir sur les modalités de sa production. Ces différents critères de jugement d'une donnée participent à la consolidation de la donnée : en développant sa compréhension de la donnée, le réutilisateur détermine sa flexibilité interprétative, c'est-à-dire sa capacité à répondre à ses besoins.

2) Le nettoyage : préparer la donnée pour un usage secondaire

La seconde opération préalable à la réutilisation des données est le nettoyage. Elle regroupe l'ensemble des tâches de préparation d'une donnée en vue d'un usage particulier. Ces opérations sont variées : identifier et corriger des erreurs de saisie, détecter et retirer les données non pertinentes, repérer les incohérences éventuelles, restructurer les données pour permettre leur insertion dans un autre modèle de données, transformer la mise en forme des données ou ajouter quelques éléments nécessaires à un usage secondaire.

Ce traitement préalable des données, dont le temps qui lui est dévolu peut être important, est toujours lié à l'usage qui sera effectué de la donnée.

Enquêteur : À chaque fois, il y a un travail assez important préparatoire sur la donnée avant de pouvoir la rentrer dans votre modèle ?

Manon : Oui. Oui, oui. (Affirmative) Il y a toute la dimension traitement de la donnée, il y a toujours besoin de retraiter la donnée, de la reconsolider, de la fusionner pour pouvoir la rentrer dans les modèles. Cette dimension traitement va être aussi automatisée dans une certaine mesure, après, on ne peut jamais tout automatiser [...] Encore une fois on revient sur l'usage et le besoin, c'est-à-dire que de toute façon, le traitement est toujours adapté en fonction du besoin, de l'usage qui en est fait à terme².

¹ Ixchel M. Faniel et Trond E. Jacobsen, « Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data », *Computer Supported Cooperative Work*, 2010, vol. 19, n° 3, p. 355-375.

² Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

Le « nettoyage » peut consister à corriger les erreurs de saisie, qui empêchent le traitement automatique des données, comme le souligne Alice, responsable SIG au sein de la municipalité de Villeurbanne.

L'homogénéité des jeux de données est parfois incertaine dans leur saisie, ce qui pose des problèmes en cas de traitement automatique. Par exemple, pour la donnée sur les marchés, j'ai des noms qui ne sont pas écrits de façon identiques, j'ai d'un côté Avenue, de l'autre Ave, au niveau de la structure du nom, c'est également différent, pour certains, il y a « espace / espace », pour d'autres, pas d'espace. C'est pas très important pour lire la donnée comme ça, mais ça pose des problèmes quand on veut faire des traitements automatiques, et retirer par exemple tous les 15 premiers caractères¹.

nom	commune	surface
VILLEURBANNE/Ave Roberto Rossellini	VILLEURBANNE	907
VILLEURBANNE / Place Victor Balland	VILLEURBANNE	3211
VILLEURBANNE / Avenue Saint Exupry	VILLEURBANNE	1475
VILLEURBANNE / Place Croix-Luizet	VILLEURBANNE	1712
VILLEURBANNE / Place Chanoine Boursier	VILLEURBANNE	6115
VILLEURBANNE / Place Wilson_Di	VILLEURBANNE	18682
VILLEURBANNE / Rue Pierre-Joseph Proudhon	VILLEURBANNE	1580
VILLEURBANNE/ Place Wilson Sem	VILLEURBANNE	5852
VILLEURBANNE / Place Grandclmnet	VILLEURBANNE	11563
VILLEURBANNE/Place de la Paix	VILLEURBANNE	1101

Illustration 27 : Extrait de la donnée « Marchés forains du Grand Lyon »

Nettoyer la donnée peut également signifier ajouter des informations complémentaires. Par exemple, le responsable SIG de la ville de Lyon insère quelques adresses non présentes dans le fond de plan fourni par le Grand Lyon afin que les internautes puissent localiser certains lieux (Opéra, hôtel de ville, parcs et jardins, etc.). Ces modifications, marginales, permettent un nouvel usage du jeu de données.

Le fond de plan, on le modifie. On l'enrichit. Notamment sur la base voies, on modifie un peu, parce que ça ne convient pas tout à fait [à notre usage]. [...] Par exemple, la ville de Lyon décide parfois de dénommer des voies ou des espaces, typiquement des squares ou des jardins. Or, ces squares ne sont pas reconnus comme des voies en tant que telles par le Grand Lyon, donc, ils ne seront jamais créés dans la base voies du Grand Lyon. Mais pour nous, à la ville, ça existe. Par exemple, le jardin des Enfants d'Izieu, c'est un petit jardin au milieu de deux rues. Mais pour le Grand Lyon, ce n'est pas une voie, donc ce n'est pas dans la base voies. Donc, nous on les rajoute ces voies, parce que nos élus ont voté ça, donc, pour les élus de la ville de Lyon, ces voies existent.

¹ Source : Entretien avec Alice, responsable SIG, mairie de Villeurbanne (7 août 2014)

[...] On a ce cas de figure assez fréquemment : on crée des numéros complémentaires pour enrichir la base¹.

Ces derniers cas présentent des situations d'arrangement de l'utilisateur avec la donnée. Puisqu'il ne peut faire exactement ce qu'il souhaitait avec la donnée, l'utilisateur s'adapte en ajoutant quelques éléments afin de contourner les contraintes de la donnée. L'enrichissement de cette donnée la dote d'une flexibilité interprétative lui permettant d'être utilisée au sein d'un autre cadre d'action sans perdre sa signification initiale. Cette opération de consolidation de la donnée est réalisée dans cet exemple par le réutilisateur. Toutefois, elle pourrait être effectuée par le producteur, qui, prenant en compte ces usages, ajouterait des informations aux données dès leur conception. Quelques administrateurs de données du Grand Lyon s'interrogent en ce sens.

Doit-on mettre les données brutes sur la plateforme ou faut-il les travailler pour faciliter leur réutilisation ? On se pose cette question car quand on voit les résultats des autres initiatives open data, on voit qu'ils ne sont pas très éloquents. On a discuté avec Veolia qui nous disait qu'ils avaient passé six mois avant de pouvoir utiliser les données mises à disposition par Rennes. Donc, n'y a-t-il pas moyen pour nous de faciliter la réutilisation ? Par exemple, aujourd'hui, on diffuse des données sur les voies, mais ces données ne contiennent pas le graphe de circulation (sens unique...). Les entreprises sont obligées d'acheter ces données à d'autres entreprises, type Navteq ou autres. On réfléchit donc, et ce n'est pas si simple, pour intégrer ces informations dans nos jeux de données².

Qu'est ce que l'ouverture des données change ? Un administrateur de données, tout ce qu'il fait, les renseignements, les attributs de données, tout est tourné vers un service technique. Donc, on aura pour une donnée type tout ce que ça implique en terme de service, de travail pour son équipe, mais pas du tout tourné vers un usager. Déjà, cet œil là, c'est nouveau, ça va nous faire réfléchir sur la donnée en elle-même, comment on va pouvoir l'étoffer pour qu'elle soit réutilisable, intéressante pour le développement d'un service public. Pour l'instant, je pense que personne ne produit de la donnée avec cet œil là³.

Afin de développer la flexibilité interprétative des données et ainsi faciliter leur usage au sein d'autres mondes sociaux, les producteurs réfléchissent à enrichir leurs données. Toutefois, la prise en charge de cette opération de consolidation les obligerait à modifier leurs pratiques et leurs outils pour prendre en compte des intérêts externes à ceux de leur organisation. Il faudrait alors repenser la chaîne de production des données en modifiant les pratiques, les normes, les infrastructures et les processus existants pour prendre en compte un

¹ Source : Entretien avec Philippe, administrateur de données, Ville de Lyon (25 juillet 2014)

² Source : Entretien avec Nicolas, direction des systèmes d'information, Grand Lyon (16 janvier 2013)

³ Source : Entretien avec Pauline, administratrice de données, Grand Lyon (13 octobre 2013)

potentiel de réutilisation des données. Cela transformerait également le rôle des acteurs publics : il ne s'agirait plus uniquement de mettre à disposition les données utilisées quotidiennement, mais de produire des données pour des utilisateurs externes. Or, les coûts techniques, humains et financiers nécessaires pour retravailler la donnée pour des acteurs externes sont autant de ressources en moins dans la production de données pour les besoins internes.

On a rencontré NavTeq il y a deux mois, justement dans une optique de se dire : « tiens, qu'est-ce qu'il manque à la base ARIANNE pour devenir une base de données circulaire ? » Et on s'est rendu compte que c'était largement au delà de nos moyens. Ils ont 200 informations différentes sur un tronçon qui permettent de calculer tout un tas de choses ! [...] Donc, on est loin, loin, loin, avec notre base adresses, d'avoir ce niveau de calcul. On est pas capable de le faire aujourd'hui, ou alors, il faut beaucoup plus de moyens humains que ce que l'on a aujourd'hui. Derrière il faut saisir. Et on ne sera jamais au niveau d'une base de données commerciale, sachant que ces bases de données, elles s'achètent, et essayer de le faire en interne n'a pas de sens, par rapport au coût d'une telle base.¹

Ainsi, consolider une donnée pour des usages variés n'est pas une tâche aisée. Cela suppose de repenser l'ensemble de la chaîne de production de la donnée. Le rôle des acteurs publics en est également transformé puisqu'ils ne produisent plus uniquement des données pour leurs besoins internes, mais deviennent fournisseurs de données pour des acteurs externes. Cela représente un coût pouvant être difficile à assumer pour les collectivités. Faut-il effectuer ce travail substantiel si la demande de réutilisation est faible ? La réponse s'oriente plutôt vers la négative, d'autant que, comme cela est mentionné dans ce dernier extrait d'entretien, des acteurs commerciaux se positionnent sur ce rôle d'agrégation et de standardisation de données hétérogènes. La consolidation de la donnée est alors effectuée en aval de la diffusion.

*

* *

Le sourcing et le nettoyage sont deux opérations de consolidation des données ouvertes au travers desquelles leur flexibilité interprétative est mise à l'épreuve par les utilisateurs. Elles permettent de s'assurer que ces données puissent répondre à d'autres cadres d'actions sans perdre leur signification initiale. Ces opérations ne sont toutefois pas suffisantes pour que

¹ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

la donnée devienne un objet-frontière utilisée au sein d'autres mondes sociaux. Il faut également que l'infrastructure sur laquelle elle repose soit partagée entre les mondes sociaux en interaction.

C. L'infrastructure partagée : propriété essentielle des données ouvertes en tant qu'objet-frontière

Revenant sur le concept d'objet-frontière et ses utilisations diverses, Susan Leigh Star insiste sur l'attachement fort établi initialement entre objet-frontière et infrastructure. La « flexibilité interprétative » n'est pas une condition suffisante pour caractériser des objets d'objet-frontière¹. « *La flexibilité interprétative sans la compréhension de l'infrastructure, des besoins en information, en standards et en classification est une erreur d'interprétation* »². La coordination est objectivée par le partage de formes communes (normes, règles, représentations collectives) qui structurent l'interaction entre différents mondes sociaux. Les données ouvertes ne font pas exception : le partage d'une infrastructure, composée notamment de standards et de conventions d'usage, entre les producteurs et les réutilisateurs, est nécessaire pour que la donnée devienne objet-frontière. La donnée peut facilement dépasser les frontières organisationnelles si elle répond aux conventions propres à un milieu professionnel. Les cas des données géographiques et des données de mobilité soulignent que leur usage est facilité par l'utilisation de formats standardisés.

Les données du système d'information géographique du Grand Lyon sont aisément réutilisées par des géomaticiens d'autres institutions. Les communes de la communauté urbaine sont des utilisateurs importants de ces données qu'elles « ne modifient pas et prennent telles qu'elles »³. De même, l'entreprise Oslandia, spécialisée en SIG *open source*, utilise les données *open data* (données 3D, cadastres, données vectorielles en 2D, bornes Velov, etc.) du Grand Lyon pour ses projets de recherche et développement sans contrainte majeure.

¹ G.C. Bowker et S.L. Star, *Sorting Things Out. Classification and Its Consequences*, *op. cit.*

² « *Interpretive flexibility without understanding of infrastructure, information needs, standards and classification is a misunderstanding* ». Conférence de Susan Leigh Star à Grenoble en 2007, cité par

P. Trompette et D. Vinck, « Retour sur la notion d'objet-frontière », *art cit.*, p. 16.

³ Source : Entretien avec Philippe, administrateur de données, Ville de Lyon (25 juillet 2014)

Sur les fichiers, on a peu de problèmes. On est des experts du secteur, donc c'est plutôt facile pour nous. [...] Généralement, le format des données quand il s'agit d'un standard OGC ou ISO, ou un standard ouvert, ça nous convient bien. Le CityGML pour les données 3D, c'est très bien pour nous. Sinon, pour le traitement des données, on n'a pas eu de difficultés particulières, c'est notre métier. De même, en terme de compréhension de la donnée, on n'a pas eu de difficultés¹.

Chez ces acteurs dotés d'une « culture SIG », les formats proposés par l'infrastructure de diffusion de données du Grand Lyon conviennent parfaitement.

On a utilisé pas mal de leurs données. En carto, on utilise énormément le parcellaire, parce que le parcellaire cadastral est super important pour certaines problématiques que l'on traite, on a utilisé pas mal de choses en termes de 3D. On a récupéré le CityGML de tout le Grand Lyon, toute la donnée 3D, ça a été récupéré. [...] Sur les formats, ils balayent pas mal de choses, et qu'en tant que Sigiste, du moment qu'il y a du shape, on peut se débrouiller. Même pour les données 3D, il se trouve que le format qu'ils proposent est un format qui nous va bien².

Extraites du SIG du Grand Lyon et mises à disposition selon des formats normalisés par l'*Open Geospatial Consortium*, ces données s'intègrent aisément au sein d'autres systèmes d'information géographique. Leur association à ces nouveaux utilisateurs ne suscite pas de « data frictions » majeures³. De manière similaire, les données de transport en commun, disponibles sous des formats standardisés (GTFS, Neptune), sont utilisées par plusieurs entreprises qui développent des services d'informations à destination des usagers⁴.

Tant dans le secteur de l'information géographique que dans celui de la mobilité, ces formats standardisés sont autant de conventions entre des communautés professionnelles qui permettent à la donnée d'être utilisée au delà des frontières organisationnelles. Un géomaticien n'aura pas de difficultés à insérer une couche de données géographiques dans son SIG ; dès lors qu'elles sont fournies dans un format standardisé, un développeur parviendra à utiliser les données des horaires des transports en commun pour mettre en place un service d'information aux usagers du réseau. Ainsi, la mise en relation de ces groupes professionnels dépend de l'existence de tout un ensemble de conventions partagées. La capacité de la donnée

¹ Source : Entretien avec Pascal, fondateur de l'entreprise Oslandia (26 janvier 2015)

² Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

³ A contrario, des professionnels d'autres secteurs d'activité peuvent rencontrer des difficultés à utiliser ces données géographiques. L'équipe en charge de l'infrastructure de diffusion reçoit ainsi régulièrement des emails d'architectes ne réussissant pas à utiliser les formats shape ou CityGML, et qui souhaiteraient que les données soient diffusées dans un format compatible avec le logiciel Autocad.

⁴ De manière non exhaustive, ces applications disponibles sur smartphone sont : Gone, Citymapper, Moovit, Transit App, Optymod, etc.

à devenir objet-frontière résulte du partage de ces appuis conventionnels, soubassement invisible de la coordination entre mondes sociaux.

*

* *

Devenue objet-frontière, la donnée ouverte permet à différents mondes sociaux de maintenir leurs différences et leur autonomie, tout en se coordonnant dans le temps et dans l'espace. Les acteurs peuvent alors s'accorder autour d'une représentation commune de l'espace urbain tout en préservant leurs points de vue respectifs. Selon le responsable SIG de l'agence d'urbanisme de Lyon, l'*open data* assure un langage commun à l'ensemble des acteurs d'un projet urbain (collectivités, agence d'urbanisme, cabinets d'architecte, etc.).

Il y a plein de cabinets d'archi qui ne connaissaient pas l'open data, donc, dans la note, on a mis, le téléchargement du fond de plan est à tel endroit, et on leur a bien dit vous allez télécharger sur internet, parce que vous êtes sûr que vous aurez le dernier mis à jour. Au moins, on ne se pose pas de questions. C'est celui que nous on utilisera pour faire la saisie du PLU, donc, c'est sûr que l'on aura la même chose. Vous n'allez même pas essayer de contacter la DGFIP en disant je voudrai récupérer le cadastre pour faire le projet. Non, Non, c'est à tel endroit, vous le prenez à tel endroit. Et pareil, c'était de dire, vous n'appellez pas le service diffusion des données du Grand Lyon ou un autre service qui pourrait avoir les données, parce que peut-être qu'il ne va pas vous les donner dans le même système de projection ou autre, s'il n'est pas au courant, si vous lui expliquez pas la demande, non vous allez tous là [sur le site open data], et vous le prenez comme ça. [...] Tout le monde a le même fichier, ils parlent le même langage. Ils sont tous en CC46 [un système de projection], on ne se pose pas de questions, on est tranquille. On est sûr que tout le monde utilise le même référentiel, et qu'ils utilisent une donnée à jour¹.

Ainsi, en acquérant le statut d'objet-frontière par l'intermédiaire de diverses opérations de consolidation, la donnée ouverte devient un point de passage obligé et assure la coordination entre des mondes sociaux hétérogènes. La mise en place de cet espace d'équivalence renforce la capacité du Grand Lyon à imposer sa représentation de l'espace urbain à l'ensemble des acteurs intervenant sur son territoire. Toutefois, comme cela a été démontré, les données ouvertes ne sont pas en soi des objets-frontières : il existe de nombreux cas de réutilisation qui traduisent, voire trahissent, les significations initiales de la donnée, au risque pour le Grand Lyon de perdre la maîtrise de la coordination de ces mondes sociaux hétérogènes.

¹ Source : Entretien avec Olivier, responsable SIG, agence d'urbanisme de Lyon (25 février 2015)

II. Homogénéisation

L'homogénéisation constitue la deuxième politique de réutilisation des données ouvertes. Elle repose sur la production d'un nouvel agrégat à partir de données hétérogènes. Contrairement à la consolidation, il ne s'agit pas ici d'utiliser la donnée ouverte pour un nouvel usage tout en préservant sa représentation de l'espace urbain, mais de la considérer comme une source d'information parmi d'autres pour construire une nouvelle donnée uniformisant des représentations hétérogènes. L'homogénéisation est ainsi la construction d'un agrégat basé sur des conventions d'équivalence alternatives à celles établies par les producteurs des données ouvertes. L'homogénéisation repose sur un processus de transformation de données hétérogènes, produites selon des spécifications variées en une donnée normée, représentant de façon uniforme un espace¹. Ces opérations d'homogénéisation font émerger une nouvelle représentation de l'espace urbain, au sein de laquelle l'héritage institutionnel des données n'est plus visible.

Ce besoin d'homogénéisation émerge lorsqu'il y a une volonté d'utiliser les données à une échelle plus large que celles pour lesquelles elles ont été produites. Cela peut-être le souhait d'utiliser des données communales à l'échelle intercommunale, des données métropolitaines à l'échelle nationale ou internationale ou plus largement sur un périmètre autre que les territoires institutionnels. L'homogénéisation vise à dépasser la tension, pointée par Alain Desrosières, « *entre d'une part le territoire, espace d'équivalence associé à une action institutionnelle et susceptible d'être découpé, et d'autre part la localité, comme support d'une vie sociale, économique et politique aux composantes et interactions multiples, dont les habitants peuvent souhaiter des descriptions synthétiques informant des actions variées*² ». Alain Desrosières associait le territoire à l'espace national, il s'agit davantage ici des territoires des institutions du gouvernement urbain, dont le périmètre diverge des espaces locaux, nationaux ou mondiaux que souhaitent représenter certains acteurs par

¹ Dans une perspective similaire, Paul Edwards analyse magistralement comment des données environnementales mondiales ont été produites à partir de données météorologiques nationales hétérogènes. Paul Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge Mass., MIT Press, 2010.

² Alain Desrosières, « Le territoire et la localité. Deux langages statistiques. », *Politix*, 1994, n° 25, p. 46.

l'homogénéisation. En uniformisant les références au territoire institutionnel, l'homogénéisation crée une nouvelle convention d'équivalence autonome de toute action institutionnelle. La perte de cette maîtrise de représentation de l'espace par les acteurs publics reconfigure le pouvoir urbain au profit de nouveaux acteurs positionnés en situation d'intermédiaire entre producteurs et utilisateurs de données.

Ces opérations d'homogénéisation de données ouvertes hétérogènes sont étudiées à partir de quatre cas utilisant les données du Grand Lyon : la base de données géographiques OpenStreetMap (OSM), le producteur de bases de données géographiques Here, le service Google Transit et le calculateur d'itinéraires en transport en commun Navitia.io, proposé par l'entreprise Canal TP.

A. Des données ouvertes hétérogènes aux périmètres limités

Le besoin d'homogénéisation part du constat de l'hétérogénéité des données ouvertes par les institutions. Cette diversité est liée aux différents environnements informationnels au sein desquels les données ont été produites. Elle devient un enjeu lorsque des acteurs souhaitent développer des services qui dépassent le périmètre spatial et/ou institutionnel de ces données. Ils doivent alors mettre en relation les données provenant de différents producteurs et créer un espace d'équivalence qui ne soit plus celui du territoire institutionnel.

Ce travail de mise en relation des données par homogénéisation est effectué par certains acteurs qui se positionnent en situation d'intermédiaire entre les producteurs et les utilisateurs de données. Par exemple, constatant que la multiplicité des portails *open data* et l'hétérogénéité des données sont un frein à leur réutilisation, l'entreprise Canal TP a développé une plateforme qui centralise et standardise ces données, puis les met à disposition de développeurs externes. Matthieu, responsable marketing de l'entreprise, présente cette stratégie :

Quand il y a eu l'ouverture des données sur Paris pour prendre un exemple concret, la SNCF/Transilien a fait son site avec sa plateforme open data, la RATP a fait son site avec sa plateforme opendata, etc.. Et donc, les développeurs, il aurait fallu qu'ils aillent chercher les données sur chacun des sites, sachant qu'elles ne sont pas forcément construites selon les mêmes référentiels. C'est vite le bordel. [...] C'est du temps de perdu pour les développeurs. Nous le travail que l'on va faire, c'est d'agréger les différents fournisseurs de données sur un même

territoire, pour qu'ensuite le développeur n'ait pas à se poser ces questions là. Il a sa donnée, il a les bus, les trains, les tramways, etc. [...] L'ambition que tu peux mettre derrière Navitia.io, c'est d'avoir un modèle de plateforme en fait, avec plein d'acteurs transporteurs, AOT, autour de la mobilité, peut-être aussi d'autres acteurs, qui vont apporter d'autres données, qui vont intéresser, et puis une communauté de développeurs en fait, ou de start-up, qui vont réutiliser ces données. Et là, du coup, ils peuvent avoir des données open data de plein de réseaux de transports, de différentes villes, de différents pays. Parce qu'on a toute la Hollande, on a tout le Royaume-Uni¹.

La plateforme navitia.io rassemble aujourd'hui les données de transport en commun de plus de 130 opérateurs de transport aux Etats-Unis, au Canada, au Royaume-Uni, en Belgique, en France, etc².

De manière similaire, l'entreprise de cartographie numérique Here Maps produit des bases de données géographiques destinées à la navigation routière³. Leur production repose sur de nombreuses sources d'informations⁴. L'usage de l'*open data* est réservé exclusivement aux mises à jour de la base de données de Here. « *Le principal avantage de l'open data, c'est que ça nous permet une mise à jour plus rapide de nos données. Sinon, la mise à jour aurait été faite mais je ne sais pas quand* »⁵. Ses cartographies, qui couvrent près de 190 pays, sont vendues à de nombreux constructeurs automobiles⁶, mais aussi à des entreprises numériques (Amazon, Bing, Yahoo !, etc.). Pour être associées aux bases de données géographiques de l'entreprise Here, les données en *open data* doivent respecter des normes précises édictées par la société. Cela doit assurer un degré de qualité similaire sur l'ensemble des territoires couverts par la société. A partir de sources de données hétérogènes, ce travail de qualification

¹ Source : Entretien avec Matthieu Responsable marketing et communication, Canal TP (28 janvier 2015)

² La liste des jeux de données intégrés est disponible à cette adresse : <http://www.navitia.io/datasets> (Consulté le 29 août 2015)

³ Here Maps est l'héritière de la société américaine Navteq. Fondée en 1985, Navteq a été achetée 5,7 milliards d'euros en 2007 par l'entreprise Nokia. Cette dernière a revendu à l'été 2015 cette activité cartographique à un consortium de constructeurs automobiles allemands (Audi, BMW et Daimler) pour un montant de 2,8 milliards d'euros. L'entretien sur lequel s'appuie la présentation de ce cas a été réalisé avant cette vente de Here.

⁴ La couverture de la France a débuté en 2000 par l'achat d'un fond de carte à l'IGN, la BDTopo, qui a été complété par le passage de véhicules traceurs dotés d'antennes GPS sur l'ensemble des routes françaises. Depuis 2005, l'entreprise se contente de faire des mises à jour de cette base de données initiale. Pour cela, le territoire est divisé en plusieurs agences, dont une à Lyon qui couvre la zone Rhône-Alpes, Auvergne et Bourgogne. L'entreprise utilise plusieurs modalités pour actualiser sa base de données : des véhicules traceurs, de la veille internet, des partenariats avec des collectivités et des entreprises, des signaux GPS provenant des GPS embarqués, des remontées de leurs clients, et, depuis 2012, l'open data.

⁵ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

⁶ Here indique fournir des informations géographiques à quatre voitures sur cinq équipées de systèmes de navigation embarqués.

et d'intégration fournit alors une représentation homogène de l'ensemble des espaces couverts par l'entreprise¹.

OpenStreetMap (OSM) est un projet de constitution collaborative d'une base de données géographiques, placée sous licence *Creative Commons*. Surnommé le « Wikipédia de la cartographie », OSM repose sur les contributions de milliers de cartographes amateurs, qui éditent et enrichissent bénévolement des portions de cette carte collective². Les données publiées en *open data* sont utilisées par les contributeurs pour compléter cette base de données.

Le dernier cas étudié est celui de Google. L'entreprise américaine propose depuis février 2005 un service de cartographie numérique : Google Maps. Il offre des cartes, des images satellites et de rues, des informations de circulation routière et un service de calcul d'itinéraire. Google Maps reçoit plus d'un milliard d'utilisateurs par mois et son API, qui permet d'insérer ces cartes dans des sites tiers, est utilisé par plus d'un million de sites. Si elle propose une représentation unifiée de l'espace, l'entreprise utilise marginalement des données en *open data*. J'ai cependant choisi de présenter ce cas, car il est, paradoxalement, très régulièrement mentionné par les producteurs publics comme un utilisateur majeur de leurs données.

Encadré 13 : La production des cartes Google Maps

La production des cartographies de Google Maps repose sur plusieurs sources de données. L'entreprise Google s'est appuyée initialement sur des bases de données publiques³

¹ Même si certains territoires peuvent être actualisés à différents intervalles en fonction des enjeux perçus. « Le Grand Lyon reste notre priorité 1, en fait le Grand Lyon, on considère que l'on fait une mise à jour chaque mois presque. Sur Saint-Etienne, ce ne sera qu'une fois par an. En fonction des zones, il y a des priorités de mise à jour. » (Simon, analyste géographe, Nokia Here)

² Michael Goodchild, « Citizens as sensors: the world of volunteered geography », *GeoJournal*, 2007, vol. 69, n° 4, p. 211-221 ; Sarah Elwood, « Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS », *GeoJournal*, 2008, vol. 72, n° 3, p. 173-183 ; Jeremy Crampton, « Cartography: Maps 2.0 », *Progress in Human Geography*, 2009, vol. 33, n° 1, p. 91-100.

³ L'US Census Bureau aux Etats-Unis, l'IGN en France par exemple. La liste des sources de données de Google Maps est disponible à cette adresse : http://www.google.com/intl/fr/help/legalnotices_maps.html (Consulté le 7 septembre 2015)

ou privées¹. Ces données initiales ont été harmonisées, actualisées et enrichies à l'aide d'images satellites, aériennes et de rue (StreetView), traitées par des algorithmes d'analyse d'images², dont les opérations sont vérifiées et corrigées par des centaines d'opérateurs. Enfin, l'entreprise fait appel aux utilisateurs au travers de trois dispositifs : le service *MapMaker* qui permet à l'internaute d'ajouter des informations, l'onglet *Report a Problem* pour signaler une erreur et un outil interne appelé « Boule de cristal » qui analyse les réseaux sociaux et les sites internet pour identifier des sources d'erreurs. Le croisement et l'harmonisation de ces données hétérogènes reposent sur un travail manuel considérable effectué par des opérateurs internes³.

La base de données est actualisée en permanence : il n'y a pas de campagnes de mise à jour ou de *versioning* des données, les mises à jour sont continues. Les données sont standardisées pour l'ensemble des zones géographiques représentées : cette standardisation s'appuie sur des modèles de spécifications qui décrivent précisément les modalités de représentation de chaque entité.

A cette cartographie initiale, Google a ajouté deux services relatifs à l'information des voyageurs : Google Traffic et Google Transit. Google Traffic indique en temps réel les conditions de circulation routière. Il a été lancé en février 2007 dans trente villes des Etats-Unis. Il s'est maintenant généralisé. Les informations ne proviennent pas des pouvoirs publics, mais sont obtenues par les localisations GPS transmises par les utilisateurs de téléphones portables. L'entreprise calcule la vitesse de déplacement de ces utilisateurs le long d'une voie pour déterminer les conditions de déplacement⁴.

Le second service, Google Transit, fournit un calcul d'itinéraire de déplacement en transport en commun. Il a été lancé en octobre 2007 par l'intégration de données des réseaux

¹ Google achète des bases de données géographiques auprès de l'entreprise TomTom.

² Ces algorithmes permettent de détecter et d'interpréter les panneaux de signalisation routière, les noms et numéros de rue et certains points d'intérêts (enseignes commerciales, etc.).

³ L'entreprise est très discrète sur le nombre d'employés travaillant à la production de Google Maps, mais plusieurs centaines d'opérateurs sont nécessaires pour cartographier un pays, dont une large part sont basés à Bangalore en Inde.

⁴ Source: <http://googleblog.blogspot.fr/2009/08/bright-side-of-sitting-in-traffic.html> (Consulté le 5 septembre 2015).

de transports publics de plusieurs villes américaines. Pour faciliter les échanges et l'intégration de ces données de transport, Google a créé un format spécifique, le *General Transit Feed Specification* (GTFS), devenu un standard de fait. Les transports en commun sont un secteur sur lequel l'entreprise est aujourd'hui dépendante des données des opérateurs publics, puisqu'elle ne peut obtenir autrement les informations sur les horaires. Google pourrait utiliser les données mises à disposition en *open data*, pourtant, elle ne le fait pas systématiquement (voir supra). Si les données de la RATP sont intégrées au service Google Transit, ce n'est pas le cas de Rennes, Lyon, Toulouse ou Nantes par exemple. Cette absence d'utilisation des données en *open data* interpelle, alors que Google est un acteur incontournable dans le discours des producteurs suscitant de nombreuses craintes.

B. Les opérations d'homogénéisation des données ouvertes

L'homogénéisation est effectuée au travers d'un travail d'harmonisation des données publiées en *open data*, qui sont ensuite intégrées dans les bases de données de chacun de ces acteurs. Ce travail se décompose en plusieurs opérations : l'identification des données, l'analyse des licences les encadrant, le téléchargement (automatisé) et l'homogénéisation selon un modèle commun de données.

1) Identifier les données

Cette première étape d'identification est similaire à celle du *sourcing* présentée dans la partie précédente. Elle vise à identifier et estimer la fiabilité des données candidates à l'intégration au sein des bases de données développées par ces acteurs intermédiaires.

Le souhait de l'entreprise Canal TP de se positionner comme intermédiaire nécessite de centraliser et d'uniformiser l'ensemble des données de transport en commun mises à disposition en *open data*. Ces données sont identifiées de manière pragmatique au fil des annonces des opérateurs de transport, ou, par le biais d'utilisateurs du service navitia.io :

On commence à avoir des retours avec le Google Group qui est associé à navitia.io, des demandes : « est-ce que vous ne pouvez pas rajouter ça ? », « Regardez, il y a là... ». Du coup, on espère avoir de plus en plus des demandes de ce type là, parce que finalement, ça nous évite d'avoir à trop

chercher, et on sait que ce sont des choses qui sont réutilisées par la suite, puisqu'il y a vraiment des demandes spécifiques là dessus.¹

L'usage de l'*open data* par l'entreprise Here vise à actualiser ses bases de données. La personne en charge de l'intégration sélectionne les quelques jeux de données qui concernent la thématique de leur cartographie.

Y'a profusion de jeux de données, il faut vraiment faire un tri dedans, dès fois ça part dans tous les sens. Et nous, vu notre métier, on se limite aux jeux de données routières, à la voirie, aux POI [points d'intérêt]. Sur les 500 et quelques jeux de données publiés par le Grand Lyon, la plupart ne nous intéressent pas. Il n'y en a qu'une dizaine qui nous intéresse.²

L'entreprise est particulièrement attentive à la fiabilité des données qu'elle intègre à sa base de données. Les données ne sont donc pas intégrées telles qu'elles, mais elles subissent préalablement un contrôle qualité.

On n'intègre jamais une source sans l'avoir vérifiée. On croise les sources, il faut qu'il y en ait au minimum deux. On a un taux qui doit être au dessus de 96%. On prend un petit échantillon, et si sur cet échantillon, il y a 96% minimum de bonnes choses, on intègre tout. On s'est aperçu que souvent on a une base de données qui est plus à jour que les jeux de données que l'on nous donne. [...] Le plan mode doux par exemple, c'est les zones 30, j'ai voulu l'utiliser. Et je ne l'ai pas fait, parce que je me suis aperçu que nous étions plus à jour que le Grand Lyon sur les zones 30. La fiabilité n'était pas ce que je voulais, et dans ce cas là, je ne prends pas, parce qu'on a une équipe spéciale, qui peut venir dans chaque ville, qui va sur le terrain, et qui contrôle ce qui existe vraiment en ville, et ce que nous on a sur notre logiciel, sur notre fond de carte, et ensuite, ils vont dire : « sur cette zone là, vous avez un taux de réussite de 20, 30 ou 50. » Et si j'ai intégré des données qui proviennent d'une source extérieure et qu'elles sont fausses, l'équipe va me rentrer dedans. [...] C'est normal, parce que le client veut ça. Sur les vitesses, vous avez un GPS, si c'est limité vraiment à 30 à l'heure, vous avez l'icone qui va s'afficher à 30 à l'heure, si c'est en vrai 50, et que c'est marqué 30, le client va se dire, est-ce que c'est 30 ou est-ce que c'est 50 ? S'il se fait flasher, il va râler ensuite³.

Les contributeurs d'OSM réalisent un travail similaire de contrôle des données afin de s'assurer de leur précision. Elles sont également comparées aux informations existantes dans la base de données OSM. En cas de doublon, les données recueillies par les contributeurs bénévoles ont toujours la priorité sur les sources externes de données. Une fois ces étapes préliminaires réalisées, les données sont fusionnées avec la base de données OSM afin de compléter les informations déjà disponibles.

¹ Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)

² Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

³ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

2) Analyser le cadre juridique

Une fois identifiées, l'intégration des données en *open data* est loin d'être automatique. Les acteurs effectuent une analyse juridique des licences encadrant l'utilisation de ces données afin de vérifier leur compatibilité avec celles de leur propre base de données. Or, les données sont mises à disposition sous des licences diverses en fonction des territoires (ODbL, Licence ouverte, licences spécifiques, etc.), et la question de la compatibilité de ces licences entre-elles est parfois un casse-tête juridique. « *Dès qu'on commence à combiner territoires et modes de transport, on ouvre la boîte de Pandore de l'open data avec la question des licences, de leur incompatibilité, de la réciprocité, du share-alike, etc.*¹ » Cette attention particulière au cadre juridique est indispensable pour assurer une sécurité juridique aux utilisateurs de cette base de donnée uniformisée. Ainsi, au sein d'OSM, les contributeurs vérifient la compatibilité de ces licences avec la licence ODbL utilisée par l'association.

On est extrêmement rigoureux au niveau d'OSM sur les sources de données. Ce n'est pas pour nous, c'est pour nos réutilisateurs. On ne veut pas les mettre dans une insécurité juridique par l'utilisation de nos données. C'est pour ça que l'on est à cheval sur les licences, que l'on demande une confirmation écrite. On ne fait pas n'importe quoi².

De manière similaire, l'entreprise Here s'assure de son droit à modifier, traiter, intégrer et revendre les données obtenues. Elle vérifie que le fournisseur de la donnée détienne les droits sur les données. « *Le client privé veut vraiment que l'on soit sûr des données que l'on a intégrées. Clair et net : que ça soit légal, que ça soit des données acquises, que l'on ait tous les droits dessus*³. » Les données *open data* sous licences ODbL ne sont pas utilisées par l'entreprise Here, car la clause de partage à l'identique empêche la revente des données enrichies. « *A l'époque les données étaient en licence ODBL, donc impossibilité de les utiliser car interdiction de sous-licence. Et comme on vend ce que l'on fait, ce n'était pas jouable.*⁴ »

¹ Source : Interview de Guillaume Crougneau, CEO de Canal TP, « Bâtir un modèle de plateforme autour des données de la mobilité » (27 juin 2014), <http://www.withoutmodel.com/chloe-bonnet/navitia-io-la-plateforme-de-la-mobilite-personnalisee> (Consulté le 1^{er} juillet 2014).

² Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

³ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

⁴ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

La première étape de l'homogénéisation des données ouvertes au sein d'une base de donnée unifiée est ainsi l'analyse juridique. A partir de données publiées sous des licences multiples, il est nécessaire de parvenir à rendre compatibles ces différents cadres juridiques. Certaines données peuvent alors ne pas être utilisées.

3) Accéder aux données (et les actualiser)

L'étape suivante est le téléchargement des données. Lorsqu'il s'agit de mise à jour ponctuelle d'une base de données, celui-ci ne pose pas de problème particulier. L'enjeu est plus important lorsque ces réutilisateurs souhaitent automatiser le traitement des mises à jour de données afin de s'assurer de la représentativité du territoire. La question de l'actualisation est un enjeu tant pour les producteurs de données que pour le réutilisateur, comme le témoignent ces deux acteurs. Dans le premier cas, le producteur regrette l'absence de prise en compte de la mise à jour de sa donnée dans les bases de données commerciales utilisées par les GPS. Dans le second cas, le réutilisateur déplore que le producteur n'ait pas mis à disposition la dernière version de sa base de données, rendant son service inutilisable.

Plusieurs fois, on a été sollicité : par un livreur qui trouve pas la bonne adresse, par un citoyen qui dit, « les gens ne me trouvent pas car mon adresse n'est pas dans les GPS ». Donc, là, à chaque fois, j'ai une réponse type qui dit que l'adresse est à jour chez moi et que pour qu'elle soit dans le GPS, il faut que le fabricant de bases de données, Navteq ou TeleAtlas, ait l'info, qu'il l'intègre dans sa base de données commerciale, que TomTom ou Garmin ou je sais pas quoi leur achètent une base à jour, ce qui n'est pas une obligation, et en dernier ressort, que le gugusse qui a le TomTom dans la voiture, il ait fait la mise à jour de la base de données. Alors, moi, j'ai beau être à jour, avant que le numéro arrive dans le TomTom du Mr tout le monde, il peut se passer plein de choses sur lesquelles on a aucune prise.¹

On est aussi tributaire des producteurs. Si le fournisseur de données ne met pas à jour ses données, on ne peut pas les mettre à jour à sa place. Et ça, ça peut arriver. On a eu pas mal de problèmes avec [un opérateur] l'an passé. [...] Ils ont lancé leur plateforme open data, c'était je crois en juin, et ils avaient déjà mis les données du mois de mai. Déjà, c'était mal barré. Ça se périmait hyper vite les données, parce qu'il y a toujours des nouvelles lignes de tramways, des travaux, etc. Et je crois qu'en janvier, ils n'avaient pas actualisé depuis pareil quelques mois les jeux de données, donc c'était n'importe quoi².

Ces témoignages soulignent les risques de mauvaise succession des étapes dans la chaîne de la donnée. Ces situations sont caractéristiques d'une fragilisation du nouvel agrégat.

¹ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

² Source : Entretien avec Matthieu, Responsable marketing et communication, Canal TP (28 janvier 2015)

« La fragilisation est un affaiblissement des liens entre l'agrégat et l'une de ses anciennes composantes qui aboutit à l'exclusion de celle-ci ou de plusieurs d'entre elles, et augmente le risque de rupture de l'ensemble¹ ». L'agrégat peut toujours être utilisé, mais le risque est alors que la carte ne corresponde plus au territoire.

Pour assurer la consistance du nouvel agrégat, il est nécessaire de renforcer le degré de solidité des liens entre les données ouvertes et ces nouveaux agrégats. Les réutilisateurs doivent pour cela interconnecter leurs systèmes informatiques à ceux des producteurs de données ouvertes. Cet enjeu est particulièrement prégnant pour l'entreprise Canal TP. Les données de transport en commun sont très régulièrement mises à jour par les opérateurs. Il est alors nécessaire d'intégrer ces données actualisées dans le calculateur navitia.io. Une solution serait de mettre en œuvre des chaînes automatiques de traitement qui viendraient récupérer les données sur les différents portails *open data*. Le problème est que les modalités d'accès aux données varient selon les sites : certains sont aisément accessibles, avec un lien fixe (*permalink*), certains nécessitent une connexion privée, d'autres enfin changent le nom du fichier d'une mise à jour à l'autre².

Une des principales difficultés que l'on a la stabilité dans les modes de fourniture et de structuration de données. On a des formats de données qui évoluent, et à chaque fois, on est obligé de faire de nouveaux développements. Un changement de format, de conditions d'accès, d'URL du fichier, etc. Pour certains fournisseurs de données, les changements sont même très fréquents. Il y a beaucoup d'incertitudes à ce sujet, puisque tout est nouveau pour l'ensemble des producteurs, mais on a des choses bizarres, comme par exemple avec la SNCF, qui nous fournit dans un cas un webservice, et sur un autre territoire, des fichiers plats. Ce n'est pas très grave, mais à chaque fois, il faut que l'on adapte nos processus³.

TCL n'arrête pas de changer de format, de chemin ou de structure de ses données. Ça fait planter tout le système. On a par exemple, un fichier qui s'appelle arrêts.csv, le lendemain, ils l'appellent arrêts100714.csv. Ils n'arrêtent pas de changer, et nous derrière, on a une alerte, on soulève le capot, et on voit que c'est un truc à la con comme ça, mais il faut qu'on reprenne tous nos chemins⁴.

¹ E. Didier, « La consistance du futur. La prévision statistique aux États-Unis et la Grande Dépression », art cit, p. 82.

² Ces exemples soulignent que les producteurs doivent intégrer le fait que la chaîne de leurs données s'est allongée : un changement en amont peut avoir des effets sur tous les acteurs en aval.

³ Source : Entretien avec Alexis, ingénieur Cityway (9 juillet 2014)

⁴ Source : Journal de terrain, discussion avec Harald (11 juillet 2014)

« Quand on a deux bases à mettre à jour, ça va, mais là, nous, on a entre 130 et 180 bases, ça commence à faire beaucoup s'il faut les faire à la main une à une. C'est un peu le problème¹. » La récupération des données à intervalles réguliers est ainsi une tâche particulièrement chronophage en l'absence de pérennité des modalités d'accès à la donnée. Le membre d'OSM, responsable de la base adresses nationale, est confronté aux mêmes problématiques de changement des modes d'accès aux données empêchant toute automatisation des mises à jour.

Il y a des fichiers qui sont mis à jour en live, parce que c'est un service WFS qui est derrière, et en fait, tous les jours, on peut les interroger, tous les jours on peut les mettre à jour. Et il y en a d'autres, c'est publié de façon ponctuelle. Et le problème, c'est que ce n'est pas forcément totalement automatisable. Quand vous avez une collectivité qui publie un fichier, mais que son URL change à chaque fois, pfff [il souffle], c'est un peu galère. Il n'y a pas de web services qui permettent d'interroger et puis d'obtenir les données. Vu que l'on n'a pas de point d'entrée stable, là, je suis obligé de le faire manuellement, donc je le fais, tous les trois mois à peu près, je passe en revue mes 20, 25 sources en open data, je les mets à jour².

Ce point souligne l'importance de l'infrastructure de mise à disposition des données. De nombreux portails *open data* voient leur fréquentation diminuer parce qu'ils ne s'appuient pas sur une architecture assurant la réutilisabilité pérenne des données³. Une analyse des emails reçus par l'équipe en charge du site *open data* du Grand Lyon révèle que le principal problème rencontré par les réutilisateurs est relatif à l'accès aux données (téléchargement, etc.)⁴. Ces utilisateurs sont particulièrement nombreux à se signaler lors des bugs de l'infrastructure. En réponse, l'équipe en charge de celle-ci tente d'éviter ces désagréments en indiquant par le biais de newsletters ou de bandeaux sur la plateforme les interruptions techniques programmées sur la plateforme.

¹ Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)

² Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

³ Natalie Helbig et al., « The dynamics of opening government data », *Center for Technology in Government*, 2012, p. 34.

⁴ L'analyse repose sur 283 emails reçus du 28 mai 2013 au 7 octobre 2015, à l'exception de la période du 16 octobre 2014 au 31 décembre 2014.

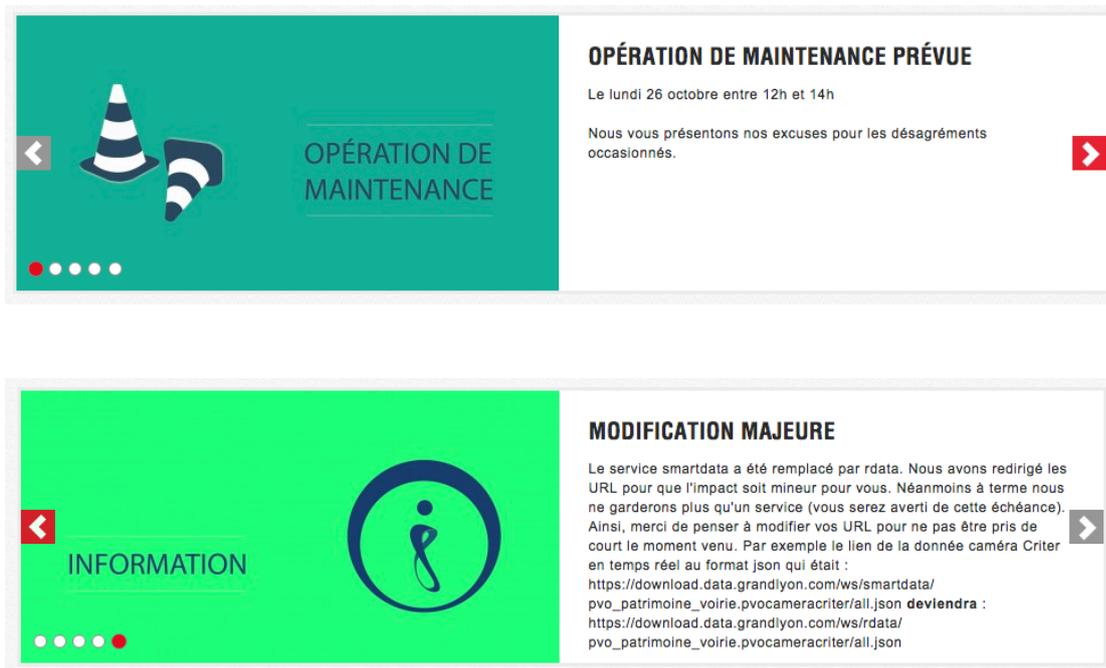


Illustration 28 : Avertissements de maintenance affichés sur le site data.grandlyon.com

4) Intégrer (et homogénéiser) les données

La dernière étape consiste à intégrer ces données ouvertes hétérogènes au sein de la base de données unifiée. Ce travail est toujours réalisé de manière semi-automatique afin de s'assurer de la compatibilité de ces deux infrastructures informationnelles. Les cas de l'entreprise Canal TP et de l'association OSM soulignent la diversité des traitements réalisés et des frictions rencontrées pour parvenir à produire une base de données unifiée.

(a) La diversité des bases de données, l'impossible automatisation de l'homogénéisation

L'homogénéisation des données est la principale difficulté rencontrée par l'entreprise Canal TP. Les opérateurs de transport sont multiples, leurs façons de produire les données le sont tout autant. « Or, produire des services homogènes à partir de structures de données hétérogènes est loin d'être simple¹. » Un des enjeux est notamment d'homogénéiser les chartes de nommage des points d'arrêts qui diffèrent d'un producteur à un autre.

¹ Source : Interview de Guillaume Crouigneau, CEO de Canal TP, « Bâtir un modèle de plateforme autour des données de la mobilité » (27 juin 2014), <http://www.withoutmodel.com/chloe-bonnet/navitia-io-la-plateforme-de-la-mobilite-personnalisee> (Consulté le 1^{er} juillet 2014).

Sur Paris, la Gare de Lyon, l'arrêt du RER, il va avoir le code 1037 à la SNCF, et y'a un arrêt de bus qui a un code complètement différent côté RATP. Potentiellement, il va avoir le code 1199, alors que dans la base SNCF, 1199, c'est la gare d'Asnières par exemple. Il y a tout un travail de remise en concordance de bases différentes sur une seule et même base, dans les endroits où les réseaux s'entrecroisent, c'est là, où c'est intéressant justement, puisque c'est ce qui permet de faire le maillage complet. Mais, la plus grosse difficulté, c'est ce problème de pérenniser des codes. Nous, faire le travail une fois, ça se comprend, c'est juste que de le refaire à chaque *data update*, c'est compliqué¹.

Pour automatiser ces traitements, Canal TP est dépendante de la pérennité de la structure des bases de données fournies par les opérateurs. Or, pour des raisons diverses, des changements importants peuvent advenir d'une mise à jour à l'autre, rendant caduque le travail précédemment effectué².

Le problème que l'on va avoir, c'est justement sur la pérennité de ces codes externes qui nous servent d'identifiants uniques. On ne maîtrise pas du tout les bases en amont. Par exemple, la RATP, on a pas encore vraiment compris comment ils fonctionnent mais leur code change régulièrement, et du coup, comme nous on crée des fusions entre les gares SNCF et les gares RATP, qui sont au même endroit, régulièrement, à chaque mise à jour, les fusions qui sont créées sautent, du coup, les grosses fusions, telles que Chatelet, où on a au moins dix zones d'arrêts, que nous on rassemble dans une seule et même zone d'arrêt, c'est un peu des problèmes que l'on a parce qu'il suffit qu'il y a un code unique qui saute sur les dix, du coup, nous on a une erreur dans la règle, et elle ne se réapplique plus³.

Enfin, l'entreprise Canal TP est tributaire de la qualité des données fournies par les transporteurs. Les opérations d'intégration et d'homogénéisation sont effectuées à Paris par une personne qui ne connaît pas les réseaux de transport de l'ensemble des villes qu'il ajoute à la base de données. En cas d'imprécisions du jeu de données initial, ou de fusions malencontreuses, il ne possède aucun moyen de corriger ces défauts de fiabilité.

Sur Paris, ça va, on connaît bien. Ça devient très vite très compliqué sur New York par exemple, où il y a des zones d'arrêt qui ont le même nom, mais qui sont à trois kilomètres d'écart. Ils sont sur la même avenue, mais ils ne sont pas au même corner. Et du coup, c'est la 1st, mais c'est au nord, c'est au sud. Et là, c'est pas beaucoup plus compliqué de faire des fusions, mais d'être sûr que la fusion soit bonne. [...] Sur un autre jeu de données, il manquait des géocodages, mais là c'est pareil, malheureusement on ne peut pas faire grand chose. Des arrêts qui sont géocodés à 0;0, et qui se retrouvent au milieu de l'Atlantique, mais nous on a aucun moyen de savoir quels sont les vrais géocodages. On a pas de contact, et on ne connaît pas en plus. Moi, j'ai aucune idée de comment est le bus à Los Angeles. Je peux sortir une carte, mais c'est pas pour ça que ça me

¹ Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)

² Encore une fois, cet exemple souligne le changement culturel qu'implique l'ouverture des données pour les producteurs. Ceux-ci doivent prendre conscience de que toute modification de leur part impacte une série plus large d'acteurs en aval de la chaîne de la donnée.

³ Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)

donnera les fréquences de passage ou ce genre de choses. On peut toujours, on sort une carte, on dit je vais de là à là, je vérifie qu'a priori, il doit prendre ce chemin là. Mais, c'est vrai qu'on n'a pas de vision approfondie du terrain. C'est un peu le problème¹.

En l'absence de connaissance exhaustive, cet intégrateur de données s'appuie sur des approximations qui lui permettent de disposer d'un jeu de données, dont la fiabilité est suffisante pour fournir un service de calcul d'itinéraire aux voyageurs. « *On essaye de maîtriser au maximum les effets que ça peut avoir. Le géocodage, c'est juste que l'on ne pourra pas faire de calcul d'itinéraire en partant d'une adresse près de ce point, mais après, on sait que ça n'a pas d'impact sur le calcul d'itinéraire derrière. Donc, ce n'est pas ça qui va stopper. Ça dégrade un peu le moteur, mais en soit, c'est mieux que rien*². » Cette déclaration souligne le changement de paradigme entre une quantification soucieuse de la précision, la représentativité et la fiabilité de la donnée, et ces acteurs, pour qui l'approximation est suffisante, dès lors que la donnée, qui s'inscrit davantage dans une perspective d'action que de description, permet de répondre à un usage.

(b) OSM : le refus de l'import massif, la priorité accordée aux contributeurs

Le groupe local des contributeurs d'OSM a utilisé certaines données mises à disposition par le Grand Lyon pour enrichir la base de données. Les orthophotographies ont permis de dessiner les contours de zones ; des couches de données diverses ont permis d'ajouter des points d'intérêts (réseaux de transport en commun, stations Vélo'v et d'autopartage, toilettes, défibrillateurs, arbres d'alignement, etc.). Lors de la publication de nouvelles données en *open data*, les membres « historiques » tentent d'éviter que des « nouveaux venus » intègrent de manière massive les données publiées en open data, remplaçant alors ce qui existe dans la base.

Il faut que l'on fasse gaffe que quelqu'un ne vienne pas bousiller tout ce qu'on a fait³.

J'espère qu'on va pas avoir des "cartographes en fauteuil" qui vont virer ce qui existe déjà et importer massivement ces données sans vérification :-(
J'avais justement inscrit un point dans l'ordre du jour de la rencontre de mardi prochain (en

¹ Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)

² Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)

³ Source : Journal de terrain, Participation à une réunion mensuelle du groupe local d'OSM (13 février 2015).

prévision de l'ouverture de ces données et suite à un différend que j'ai eu avec un contributeur), pour qu'on harmonise entre-nous notre manière de mapper le réseau TC sur Lyon...¹

En effet, avant de les intégrer, ces contributeurs comparent les données ouvertes avec les informations déjà présentes dans la base OSM. Ce travail est majoritairement manuel. Il est effectué par les membres selon leur propre envie, sur la zone qui les intéresse. Un des membres me raconte par exemple qu'il souhaitait cartographier les aires d'accueil des gens du voyage. Pour cela, il a utilisé le fichier mis à disposition par le Grand Lyon pour identifier les aires. Cependant, la donnée du Grand Lyon ne les représente que sous forme de points. Il a donc complété cette information en utilisant les orthophotographies du Grand Lyon pour tracer les polygones indiquant les aires (contour, périmètre, emplacements). Un autre membre s'est intéressé aux arbres d'alignement. Il n'a pas injecté massivement la donnée du Grand Lyon dans la base de données OSM. Il s'est en effet rendu compte que la précision de la géolocalisation était plus précise au sein d'OSM que dans la donnée du Grand Lyon. Il s'est alors contenté d'intégrer des informations spécifiques supplémentaires (circonférence, etc.) pour compléter le modèle de données OSM des informations.

OpenStreetMap joue un rôle d'homogénéisation des données en proposant une représentation uniforme d'une même entité à partir de sources de données hétérogènes. Par exemple, face à la pluralité des fichiers d'adresses existants, la branche française de l'association s'est lancée dans le défi de constituer la première base nationale d'adresses (BANO).

Le problème des bases adresses en France, ce n'est pas qu'il n'y en a pas, c'est qu'il y en a trop. C'est qu'il y en a trop et qu'il n'y en a aucune qui soit arrivée à un niveau de qualité qui fasse que l'on ne se pose même plus la question de savoir laquelle on utilise².

Pour cela, elle s'appuie sur trois types de données provenant des contributeurs OpenStreetMap, des collectivités et de l'Etat. Elle effectue un important travail d'identification des doublons entre ces données, en privilégiant toujours les données présentes dans la base OSM.

¹ Source : Echanges sur la mailing-list du groupe local d'OSM à Lyon (7 février 2015)

² Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

L'ordre de priorité, c'est :

1. On prend une adresse si elle existe dans OSM, on prend celle là, parce qu'on a pu la corriger. Elle peut être de moins bonne qualité aussi que l'open data. Donc, là, c'est pas parfait comme choix.
2. Les adresses en open data.
3. Et en troisième choix, par défaut, les données du cadastre.¹

L'objectif est de réussir à géolocaliser l'ensemble des adresses existantes et de fournir une position générique pour chacune d'entre elle. A partir de données aux structures très hétérogènes, les membres d'OSM produisent une base de données adresses à la structure homogène.

On a un peu de tout. Ca aussi, c'est un problème, parce que donc, 36 000 communes, il n'y a pas un format standardisé des données adresses. C'est l'anarchie complète. Là, j'ai une vingtaine de collectivités qui publient des données en open data, il n'y en a pas deux qui sont avec la même structure de données ! Donc, à chaque fois que je découvre une collectivité avec des adresses en open data, je télécharge le fichier, je regarde comment il est foutu, et j'écris un script pour chacune qui va s'occuper de télécharger le fichier, le décompresser, l'importer dans une base, dans une table temporaire, le réorganiser pour le mettre dans ma structure à moi, commune, voilà. Parce que on a de tout quoi ! [...], il n'y en a pas deux qui sont décrits de la même façon. C'est exactement le même problème. C'est là où OSM et BANO sont utiles, parce qu'en fait, on s'abstrait de ça vu qu'en fait on met tout dans une structure qui est la même, après si je veux des adresses, n'importe où sur le territoire, elles sont structurées de la même façon.²

En harmonisant la représentation des adresses, OpenStreetMap est un vecteur d'homogénéisation. Le résultat de cette initiative est la publication de la première base nationale d'adresses géoréférencées, qui vise à devenir une référence commune. « *A partir d'une base ouverte, on oblige les acteurs publics à mettre en commun alors que ça fait 15 ans qu'ils n'arrivent pas à se mettre d'accord pour produire une base*³. » A partir d'une pluralité de sources de données, OSM crée une convention d'équivalence de l'adresse à l'échelle nationale là où ni La Poste, ni l'IGN n'avaient réussi à construire une base nationale reconnue.

*

**

¹ Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

² Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

³ Source : Entretien avec Arnaud, contributeur OSM, responsable de la BANO (4 mars 2015)

Ces opérations d'intégration des données ouvertes au sein de ces bases de données nécessitent un travail important réalisé par des personnes dédiées au sein de chacune de ces organisations. Le travail de ces « petites mains¹ » est indispensable pour fluidifier les nombreuses frictions² qui apparaissent lors de la mise en relation d'infrastructures informationnelles. Une personne de l'entreprise Canal TP est spécifiquement en charge de l'intégration des données de transport en commun au sein du service navitia.io. De même, l'actualisation des cartographies Here est réalisée par une équipe en charge de la région Rhône-Alpes-Auvergne. L'intégration des données *open data* au sein d'OpenStreetMap repose sur un important travail manuel réalisé par les contributeurs bénévoles. L'utilisation des données ouvertes n'est pas automatisée, mais elle nécessite ainsi une importante infrastructure humaine pour permettre leur intégration au sein d'un nouvel environnement informationnel. Sans ces « petites mains », les maillons de la chaîne de la donnée ne pourrait être assemblés, et le réseau sociotechnique ne « tiendrait » pas.

C. Google : une politique alternative d'homogénéisation des données publiques

Plus les sources de données sont nombreuses et hétérogènes, plus ce travail d'intégration est particulièrement long et coûteux. Afin de ne pas le prendre en charge, l'entreprise Google a décidé d'avoir recours de manière marginale à l'*open data* et de privilégier des « partenariats » avec les producteurs faisant reposer sur ces derniers ce travail d'intégration.

Pour Transit, son service de calcul d'itinéraire en transport en commun, plutôt qu'utiliser les données *open data*, Google préfère établir des « partenariats » avec les autorités en charge des transports en commun. La première raison avancée pour justifier ces modalités d'association est l'amélioration de la qualité des informations en profitant de la connaissance des autorités locales et en leur laissant la responsabilité des mises à jour.

Pour Transit, on ne peut pas prendre les données et les injecter directement dans Transit. Il faut que ce soit la collectivité qui le fasse. Il y a une interface à laquelle la collectivité a accès, avec un

¹ J. Denis et D. Pontille, « Travailleurs de l'écrit, matières de l'information », art cit.

² P. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.*, op. cit.

compte réservé, dans laquelle elle charge ses données en GTFS. Il faut que ce soit du GTFS, mais bon, aujourd'hui, c'est un standard de fait. Ça a été développé par Google, mais maintenant le G signifie general. Ensuite, l'interface permet de visualiser les données : comment seront représentés les arrêts, les lignes avec quelle couleur, etc. Puis, si tout est ok, la collectivité valide pour l'envoi et la publication sur Google Transit. C'est vraiment un outil que l'on met à la disposition des collectivités. C'est pour ça que l'on préfère un partenariat. Ça permet d'améliorer la qualité des informations que l'on diffuse, puisqu'il y a une vérification par les acteurs locaux. Ça nous permet également d'assurer une certaine stabilité de la fourniture du service dans le temps. Par exemple, si la ville de Besançon publie ses données, mais, du jour au lendemain, change l'URL de son service, nous, on ne le verra pas forcément. Si la collectivité met à jour l'URL, ça permet d'éviter des problèmes de mise à jour, et comme ça, nous on n'indique pas des infos foireuses. C'est pourquoi, on préfère faire des collaborations avec les collectivités. C'est comme cela que l'on procède partout dans le monde, on ne vient pas utiliser directement les données en open data. On signe avec les collectivités un court contrat de mise à disposition des données et d'autorisation de les réutiliser. Puis, on ouvre un compte aux personnes dans la collectivité¹.

En signant des « partenariats », Google souhaite s'assurer d'une qualité optimale des informations fournies en s'appuyant sur l'expertise et la connaissance du réseau des acteurs locaux. Le contre-exemple de la RATP révèle en effet que le service d'itinéraires à partir de l'utilisation des données en *open data* par Google est de mauvaise qualité : seuls les métros et les RER sont intégrés, les tramways et les bus sont absents du service, qui ne prend pas en compte les perturbations du réseau². A contrario, le partenariat permet à l'entreprise d'avoir un interlocuteur au sein de la collectivité avec lequel ils peuvent travailler pour améliorer les données et le flux afin d'obtenir des « *jeux de données qui répondent à [leurs] standards de qualité* »³. Pour cela, Google donne un accès à une interface de visualisation (*Partner Dash*) des données afin d'identifier des erreurs et des corrections à effectuer. Ainsi, contrairement à l'*open data* qui est perçu comme « *plus complexe et moins engageant* », Google souhaite ajouter des médiations supplémentaires pour « *cadrer la relation* » entre les producteurs de données et l'entreprise réutilisatrice.

Ces informations de transport en commun nous sont communiquées par les collectivités. Elles ont une interface de chargement de données, qui leur permet de vérifier des incohérences avant la publication des données, éviter par exemple qu'un arrêt se retrouve en plein milieu de l'océan atlantique. On a également des ingénieurs qui vérifient des incohérences manifestes, mais la main est laissée à la collectivité. Nous, on n'est pas capable de dire si l'arrêt de bus se trouve de tel côté de la place, ou de l'autre. Quand il y a des incohérences, des erreurs à corriger, on ne le fait pas

¹ Source : Entretien téléphonique avec Hugues, directeur des politiques publiques, Google (11 février 2015)

² Les informations lacunaires rendent le service peu efficace comme le souligne ce billet de blog. Yann Le Tilly, *Google Transit à Paris : pour les cowboys uniquement !*, <http://transid.blogspot.fr/2012/11/google-transit-paris-pour-les-cowboys.html>, 28 novembre 2012, (consulté le 24 octobre 2015).

³ Source : Entretien téléphonique avec Francesca, chef de produit, Google (9 février 2015)

tout seul dans notre coin, mais on contacte le propriétaire pour qu'il améliore son flux. Nous, on n'intervient pas sur le flux directement¹.

Surtout, par le biais de ces partenariats, Google réussit à faire reposer le travail de nettoyage, d'actualisation et d'intégration de ses données dans son service sur les autorités locales². Ces dernières doivent en effet prendre à leur charge les coûts humains de formatage des données en GTFS, d'intégration, de vérification de la qualité et de mise à jour régulière des données. Le site internet présentant les partenariats est particulièrement clair sur la responsabilité qui incombe aux producteurs de données.

Il est essentiel de fournir un flux de données d'excellente qualité. Google a créé un certain nombre d'outils Open Source afin d'aider les réseaux à vérifier la qualité de leurs flux de données.

- Feed Validator : cet outil permet de contrôler le format du flux de données. **Aucun flux de données ne doit être soumis à Google tant qu'il comporte des erreurs de formatage.**
- Schedule Viewer : cet outil permet aux réseaux de visualiser tous les itinéraires et arrêts de leur réseau sur Google Maps. Utilisez-le systématiquement pour vérifier chaque itinéraire du réseau, notamment l'emplacement et l'ordre des arrêts, la vitesse du véhicule et bien d'autres informations essentielles. Pour en savoir plus, consultez ce document.
- Test dans Google Maps : dès lors que Google accepte le flux Transports en commun, une fonctionnalité de prévisualisation privée est mise à la disposition du réseau à des fins de test. Le réseau étant le mieux placé pour vérifier l'exactitude des données, celui-ci doit tester de façon approfondie les itinéraires, notamment les plus empruntés, ainsi que les itinéraires propres aux périodes de vacances ou aux week-ends. Pour que Google publie les données, le réseau doit avoir vérifié leur qualité et signé l'accord en ligne. Le flux de données doit également avoir été contrôlé par les ingénieurs Google. [...]

En tant que détenteurs des données, c'est vous qui disposez des informations les plus complètes sur votre réseau. Google vous fait donc confiance pour fournir des informations à jour et exactes. Nous préconisons l'envoi de fichiers de données valables pendant au moins 4 semaines à compter du lancement initial. **Vous devez également effectuer une mise à jour dès lors que vous avez connaissance d'une modification des horaires, au plus tard deux semaines avant que ces nouveaux horaires ne deviennent effectifs³.**

¹ Source : Entretien téléphonique avec Francesca, chef de produit, Google (9 février 2015)

² Strasbourg et Bordeaux sont les deux premières villes à avoir signé un partenariat avec l'entreprise Google pour l'intégration de données de transport en commun. Google effectue actuellement une campagne de négociation avec de nombreuses collectivités françaises afin d'augmenter le nombre de partenariats. La liste des villes présentes sur Google Transit est disponible à cette adresse : <http://maps.google.com/landing/transit/cities/index.html> (Consulté le 7 septembre 2015).

³ Source : <http://maps.google.com/help/maps/mapcontent/transit/participate.html?hl=fr> (Consulté le 24 octobre 2015) Je souligne.

L'entreprise Google, au travers de ces services Maps, Traffic et Transit, utilise marginalement les données ouvertes en *open data*¹. Le faible usage de données publiques offre à l'entreprise une indépendance vis-à-vis des producteurs officiels de données. Elle maîtrise l'ensemble de la chaîne de la donnée, de la production à l'utilisation, et peut ainsi se passer des pouvoirs publics pour mettre en œuvre un service. Cette indépendance lui permet de fournir une représentation uniformisée, homogène et actualisée continuellement des espaces couverts par ce service. Elle ne repose pas sur des sources officielles plurielles aux représentations hétérogènes. Cette homogénéisation est également recherchée lors des cas où elle est contrainte de s'appuyer sur les données produites par des acteurs publics. Plutôt que d'utiliser des données en *open data* de transports en commun qu'elle devrait harmoniser et fiabiliser, Google signe des partenariats avec les autorités locales afin que ces dernières effectuent ce travail d'homogénéisation en formatant les données en GTFS et en utilisant un outil fourni par l'entreprise pour les intégrer et en vérifier la qualité. En conséquence, Google offre une représentation uniformisée de l'espace qui fait fi des frontières administratives et institutionnelles.

L'avantage de Google est qu'il dispose de ces données dans le monde entier, et qu'il ne s'arrête pas aux frontières administratives. Ils ne se posent pas la question de savoir si cette donnée relève d'une voirie communautaire, départementale ou nationale. Ce qui est dingue, c'est qu'ils arrivent à faire croire à une continuité sans bord ni couture. On a l'impression qu'il n'y a qu'un seul espace, et que les informations sont les mêmes d'un endroit à un autre. Alors que le niveau de zoom n'est pas équivalent si tu es dans une zone en ville ou à la campagne. Il y a beaucoup moins d'infos et c'est beaucoup moins précis dans les zones rurales. Mais, à part pour les professionnels, peu de personnes s'en rendent compte, ils ont l'impression d'une continuité².

Pour parvenir à cette représentation uniformisée de l'espace, quel que soit les territoires, Google a établi des conventions d'équivalence, que l'entreprise impose aux acteurs publics souhaitant intégrer leurs données de transport en commun au sein de son service. Ce cas de l'entreprise Google fait ainsi émerger une autre politique de composition avec les données des collectivités.

¹ La faiblesse de cette réutilisation est paradoxale quand on la met en relation avec la « peur » que suscite l'entreprise Google chez les producteurs lors du processus d'ouverture de données.

² Source : Entretien avec Pierre, ingénieur trafic urbain, Grand Lyon (9 mars 2014)

D. De nouvelles conventions d'équivalence indépendantes des acteurs publics

L'intégration de sources de données hétérogènes au sein d'une base de données unifiée impose l'établissement de conventions d'équivalence¹, instaurant une représentation uniforme de l'élément représenté. Comme l'indique Alain Desrosières, les conventions d'équivalence sont indissociables des opérations de quantification. « *La construction d'un système statistique est inséparable de celle d'espaces d'équivalences, garantissant la consistance et la permanence, tant politiques que cognitives, de ces objets voués à fournir la référence des débats. L'espace de représentativité des descriptions statistiques n'est rendu possible que par un espace de représentations mentales communes portées par un langage commun, balisé notamment par l'État et le droit*². » Néanmoins, à la différence des systèmes statistiques nationaux, l'établissement de ces conventions d'équivalence n'est plus l'apanage de l'Etat ou des acteurs publics³. Il est pris en charge par des acteurs privés qui imposent leur représentation de l'espace urbain au risque, pour les acteurs publics, d'une perte de maîtrise de leurs politiques publiques. Deux cas sont étudiés ici : la diffusion du format GTFS devenu un standard des données mobilité par Google excluant des éléments de la représentation du réseau de transport en commun ; et les divergences de catégorisation de la hiérarchie des voies entre le Grand Lyon et l'entreprise Here influant sur la politique de mobilité de l'agglomération.

1) Le GTFS : un standard imposant une nouvelle représentation des réseaux de transport en commun

Au cœur des mécanismes de coordination des marchés, les standards réduisent les coûts de transaction. En limitant les variations entre les systèmes informatiques, ils facilitent l'intégration d'une donnée dans un nouvel environnement informationnel. La particularité des standards comme dispositifs de coordination est qu'ils émergent « *au travers de la médiation des processus de marché : c'est la dynamique d'adoption des acheteurs sur un marché qui*

¹ Alain Desrosières, « Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative. », *Genèses*, 2001, vol. 2, n° 43, p. 112-127.

² A. Desrosières, « Discuter l'indiscutable. Raison statistique et espace public. », art cit, p. 131.

³ Sur le cas d'OpenStreetMap et des difficultés à faire émerger une représentation alternative, voir Chris Perkins, « Plotting practices and politics: (im)mutable narratives in OpenStreetMap », *Transactions of the Institute of British Geographers*, 2014, vol. 39, n° 2, p. 304-317.

aboutit finalement à sélectionner, parmi la diversité des alternatives possibles, un ou plusieurs standards qui subsisteront »¹. Dans le domaine de la mobilité, un standard de fait a émergé et s'est imposé pour la publication des données relatives aux transports en commun : le *General Transit Feed Specification* (GTFS).

Le GTFS définit un format commun pour les horaires et les informations géographiques associées. Le format, dont les spécifications sont décrites sur le site de Google², se compose d'une série de fichiers CSV, décrivant chacun un aspect du réseau de transport en commun (arrêts, calendriers, horaires, lignes, etc.). L'ensemble de ces fichiers est compressé dans un dossier ZIP. Ce format a été développé en partenariat entre Trimet, l'autorité organisatrice des transports en commun de l'agglomération de Portland et l'entreprise Google³. L'objectif était que les données de l'AOT soient aisément intégrables au sein de la cartographie Google Maps pour fournir un service d'informations aux voyageurs⁴.

L'établissement du GTFS comme un standard de fait impose aux producteurs de convertir leurs données dans ce format, ce qui peut nécessiter un travail important puisque les systèmes d'information n'ont pas été conçus pour le traiter⁵. Les opérations d'harmonisation sont ainsi réalisées avant la diffusion des données, ce qui facilite leur intégration dans les bases de données de l'entreprise Google, mais également dans celles de multiples acteurs qui se sont adaptés à cette standardisation.

¹ Les standards se distinguent en cela des normes, qui sont des documents écrits, déterminant des spécifications techniques, élaborés par consensus entre plusieurs acteurs, et appliqués volontairement par les acteurs.

Benoît Lelong et Alexandre Mallard, « La fabrication des normes », *Réseaux*, 2000, vol. 18, n° 102, p. 20.

² <https://developers.google.com/transit/gtfs/reference> (Consulté le 19 octobre 2015).

³ Le format GTFS présente des similarités importantes avec celui alors utilisé au sein du système d'information de TriMet. Matthew Roth, *How Google and Portland's TriMet Set the Standard for Open Transit Data*, <http://sf.streetsblog.org/2010/01/05/how-google-and-portlands-trimet-set-the-standard-for-open-transit-data>, 5 janvier 2010, (consulté le 19 octobre 2015).

⁴ Bibiana McHugh, « Pioneering open data standards: The GTFS story » dans Brett Goldstein et Lauren Dyson (eds.), *Beyond Transparency*, San Francisco, Code for America Press, 2013, p. 125-125.

⁵ S. Goëta, « The Daily Shaping of State Transparency: Emerging Standards in Open Government Data », art cit.

En créant un espace d'équivalence, le GTFS rend commensurable¹ les réseaux de transport en commun des différents opérateurs au prix d'une simplification de leur représentation. Les structures de données du GTFS sont plus rudimentaires que celles, notamment, d'autres normes européennes. Par exemple, les arrêts sont regroupés sous la notion générique de « stop » sans prendre en considération leur diversité ou leur degré d'accessibilité². La représentation commune des réseaux de transport est ainsi obtenue au prix d'une perte d'informations par rapport à la donnée initiale utilisée pour l'exploitation du réseau. Comme tout investissement de forme, cette standardisation se caractérise sous la forme d'une dialectique de la perte et du gain³ : ce que la donnée gagne en commensurabilité, elle le perd en précision d'informations. Créé pour répondre aux besoins de villes occidentales et orienté vers un usage de calcul d'itinéraire⁴, la représentation des réseaux de transport véhiculée par le format GTFS réduit la singularité des points de vue dans une catégorie standard, excluant de fait et rendant invisible tout ce qu'elle ne prend pas en compte.

2) L'effet de la hiérarchisation alternative des voies sur la politique de régulation du réseau routier

Le réseau de voies de l'agglomération lyonnaise est hiérarchisé par le Grand Lyon en fonction de différents critères (volume des flux, dimensionnement de la voirie, usages, etc.). Cette classification est un instrument majeur de régulation du réseau. Néanmoins, elle est l'objet de controverses avec les producteurs privés de données de navigation routière qui proposent des typologies alternatives. « *Nous on a 5 classes de voies. Le Grand Lyon en a 6.*

¹ Wendy Nelson Espeland et Mitchell L. Stevens, « Commensuration as a Social Process », *Annual Review of Sociology*, 1998, vol. 24, p. 313-343.

² Agence française pour l'information multimodale et la billettique (AFIMB), *L'open data dans le domaine du transport : analyse des premières initiatives et recommandations*, Paris, Ministère de l'Écologie, du Développement durable et de l'Énergie, 2013, p. 14.

³ Bruno Latour, « Le "topofil" de Boa Vista ou la référence scientifique » dans *La clef de Berlin et autres leçons d'un amateur de science*, Paris, La Découverte, 1993, .

⁴ Créé pour répondre aux besoins d'une ville nord-américaine, le format GTFS est limité pour représenter les réseaux semi-informels des villes des pays en développement. Il suppose en effet que l'information soit rassemblée par une autorité organisatrice qui contrôle le réseau et que celui-ci soit composé d'arrêts, de lignes et d'horaires formalisés. Voir par exemple, ce cas d'application du format GTFS au sein de la ville de Mexico. Emily Eros et al., « Applying the General Transit feed Specification to the Global South: Experiences in Mexico City, Mexico, and Beyond », *Transportation Research Record: Journal of the Transportation Research Board*, 2014, p. 44-52.

*Je sais que l'autre fabricant de GPS en a 8.*¹ » Dès lors, selon les bases de données, certaines voies ne sont pas catégorisées de manière identique.

L'exemple flagrant, c'est Fourvière, un grand sujet de discussion [rires], qui est embouteillé matin et soir. Nous, c'est un niveau classe 2. En fait, la classe 1, c'est les autoroutes, et juste en dessous, c'est la classe 2. Donc Fourvière est en classe 2, et le boulevard périphérique nord est en classe 1. Et quand on a rencontré les personnes du Grand Lyon, ils ont un petit peu tiqué, parce que nous, on envoie, enfin nos GPS envoient les automobilistes sur le boulevard périphérique nord dans un premier temps. Et ensuite Fourvière, alors qu'ils voulaient, si on vient du Nord, et que l'on va au Sud, ils voulaient que les automobilistes prennent la rocade Est, et on leur a expliqué par A + B avec un plan, que jusqu'à ouverture de l'A466, la future déviation au nord de Lyon, qui n'est pas ouverte encore, on est obligé de laisser Fourvière en classe 2, jusqu'à ce que cette déviation soit ouverte. Cette déviation, vu qu'elle est en dehors du Grand Lyon, le Grand Lyon, pour eux, ce n'était pas inclus dans leur base, alors que nous on a une vision qui est plus large.²

Comme le met en avant cet interlocuteur, l'entreprise Here a construit une convention d'équivalence alternative à celle de la Métropole de Lyon pour représenter la hiérarchie des voies du réseau routier. La classification de l'entreprise Here a été définie afin de prendre en compte la diversité des voies à l'échelle de la région, tandis que la communauté urbaine de Lyon a construit la sienne dans une optique de régulation du réseau à l'échelle de l'agglomération. Cette différence de représentation n'est pas neutre : elle est une force de prescription importante sur le comportement des automobilistes par l'intermédiaire de l'itinéraire mis en avant par leur GPS.

Particulièrement visibles lors de situations de congestion, ces différences de classification entraînent un report du trafic automobile par les GPS sur des voies que le Grand Lyon considère comme non adaptées. Ce cas s'est produit notamment à l'été 2013 où une série d'embouteillages a saturé le réseau urbain. Les autorités et la presse locale ont alors fait porter la responsabilité sur les producteurs privés de données, exigeant qu'ils alignent leur classification sur celle du Grand Lyon.

Cette discussion sur la hiérarchie fait suite à un gros embouteillage qu'il y a eu à Lyon en juillet 2013, où la rocade Est était vide, et de Vienne jusqu'à Villefranche, via Fourvière, c'était noir en fait. Et les autorités et les journaux locaux dont le Progrès³, ont dit : « c'est la faute des GPS ». Article en grand. Quand on a vu ça, on s'est dit pfff. Donc, on a eu des réunions avec la DIR, la

¹ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

² Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

³ En septembre 2013, le quotidien régional Le Progrès titre « Les GPS contribuent à alimenter le bouchon de Fourvière pendant l'été ».

Préfecture, le Grand Lyon, on a du expliquer et prouver que ce n'était pas forcément du uniquement à nous. [...] Mais, il [une personne du Grand Lyon] voulait que l'on change, que l'on déclasse ou que l'on surclasse certaines routes. Donc, j'ai fait des changements, je n'en ai pas fait, et du coup, je lui ai renvoyé ce que j'avais changé ou pas. Je lui ai dit, si ça je l'ai pas changé, c'est parce que nous, on considère que cette route là elle doit rester comme ça. Parce que le Grand Lyon a une certaine vision de son réseau, nous aussi, parce que l'on est au niveau Rhône Alpes, alors que le Grand Lyon, c'est le Grand Lyon¹.

Ce conflit entre l'autorité en charge de la régulation du réseau routier et ces producteurs privés de base de données de navigation routière met en évidence les répercussions des représentations alternatives de l'espace urbain sur la politique publique de la communauté urbaine de Lyon. La hiérarchisation des voies établie par l'acteur public ne fait plus convention, ce qui réduit sa capacité de régulation du réseau². Face à ces nouvelles conventions d'équivalence, le Grand Lyon ne maîtrise plus la politique de mobilité de l'agglomération.

E. Conclusion : le modèle de la « plateforme » : une stratégie d'homogénéisation

L'homogénéisation est une des modalités de réutilisation des données ouvertes qui se caractérise par la création d'une nouvelle base de données qui harmonise les données mises à disposition par les acteurs publics. L'hétérogénéité de leur format, leur licence ou leur structure est uniformisée au sein d'un nouvel agrégat. Ces opérations d'homogénéisation produisent de nouvelles représentations de l'espace qui dépassent les frontières institutionnelles et permettent de penser ensemble des objets ou des événements auparavant singuliers ou incommensurables³. Ce travail est un investissement de forme, au coût humain, technique et économique conséquent, qui assure la coordination d'acteurs à une échelle alternative aux territoires administratifs. L'homogénéisation produit en effet de nouveaux espaces d'équivalence, dont le périmètre varie selon les enjeux traités, qui s'affranchissent des

¹ Source : Entretien avec Simon, analyste géographe, Nokia Here (9 janvier 2015)

² Face à cette situation, le Grand Lyon se trouve réduit à agir sur l'infrastructure, en réclamant à l'État le déclassement de la voie, afin de pouvoir réduire drastiquement la vitesse de circulation et installer des feux de circulation, ce qui contribuerait à modifier substantiellement les calculs d'itinéraires.

Source : <http://www.rue89lyon.fr/2016/04/01/autoroute-a6-a7-quand-gerard-collomb-voulait-convaincre-les-gps-dignorer-fourviere> (Consulté le 16 mai 2016)

³ Pour un cas similaire de co-construction d'une infrastructure d'échange de données et d'une nouvelle représentation de l'espace, voir P. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.*, op. cit.

espaces institutionnels. Contrairement à la statistique étatique, le territoire national n'est ainsi plus le modèle unique d'homogénéisation des informations locales¹. En outre, si ces représentations produisent des espaces homogènes, ceux-ci ne sont pas toujours continus et contigus. Les données de transport en commun de différentes villes seront par exemple uniformisées, mais, les données sur les espaces entre ces villes ne seront pas toujours intégrées. Le modèle territorial, caractérisé par son uniformité, sa continuité et sa contiguïté, ne s'applique pas automatiquement aux agencements spatiaux proposés par ces nouveaux agrégats.

L'émergence de ces nouveaux espaces d'équivalence, mis en place par des acteurs privés, reconfigure le gouvernement de la ville. Par la perte de son monopole de représentation de l'espace urbain, l'acteur public risque de perdre la maîtrise de certaines de ces politiques publiques. Ce mode de réutilisation des données ouvertes produit en effet une configuration alternative entre producteurs et réutilisateurs des données. Il voit émerger des acteurs qui se positionnent dans une situation d'intermédiaires entre producteurs et utilisateurs de données. L'entreprise Here vend sa base de données navigable à de nombreux constructeurs automobiles ; Canal TP met à disposition son service de calcul d'itinéraire en transport en commun à des développeurs² ; et l'API de Google Maps, l'une des plus utilisées sur internet, permet d'afficher ces cartographies dans de nombreux sites et applications.

Lorsqu'elle est couplée à une volonté d'acquérir une position dominante, cette situation d'intermédiaire se rapproche des stratégies de plateforme mises en place par les entreprises majeures de l'économie numérique (Amazon, Apple, Google, Facebook, etc.)³. Celles-ci se positionnent comme des intermédiaires entre des fournisseurs de services et des utilisateurs⁴.

¹ A. Desrosières, « Le territoire et la localité. Deux langages statistiques. », art cit.

² Cette orientation vers un modèle de plateforme transforme le modèle économique de l'entreprise. Alors qu'auparavant ses clients étaient les opérateurs et autorités organisatrices des transports, leurs clients sont maintenant des entreprises qui souhaitent « intégrer la mobilité dans leur service » (Source : Entretien avec Jérôme, responsable bases de données, Canal TP (4 mars 2015)).

³ Nicolas Colin et Henri Verdier, *L'âge de la multitude. Entreprendre et gouverner après la révolution numérique.*, Paris, Armand Colin, 2012.

⁴ Stéphane Frénot et Stéphane Grumbach, « Des données à l'intermédiation, une révolution économique et politique. » dans Lisette Calderan et al. (eds.), *Big Data : nouvelles partitions de l'information*, Bruxelles, De Boeck, 2015, p. 97-120.

Elles offrent tout d'abord un service gratuit en échange de la capacité à collecter et exploiter les données personnelles de leurs utilisateurs¹. Puis, elles mettent à disposition ces informations à des acteurs tiers via des interfaces de programmation afin que de nouveaux services soient développés. Un écosystème d'applications est alors créé autour de leurs plateformes. Ces entreprises se retrouvent alors en situation « d'hypermédiation² », dont sont dépendants de nombreux « surtraitants³ ».

Le modèle de la plateforme repose sur la mise en place d'une infrastructure qui agrège et standardise les données provenant d'acteurs hétérogènes, pour ensuite les mettre à disposition d'acteurs tiers. Les entreprises du numérique collectent les données - de leurs utilisateurs mais pas uniquement, les structurent uniformément selon un modèle déterminé, puis les proposent sous forme de services de diffusion (API). Appliqué aux politiques d'ouverture de données des administrations publiques, ce modèle se traduit par une centralisation de données provenant de différents producteurs qui sont mises à disposition sous une forme homogène à des acteurs tiers. Ces processus conjoints de centralisation et d'homogénéisation des données sont les caractéristiques centrales de la plateforme comme infrastructure. Ils conduisent à recomposer les relations entre producteurs, opérateur de l'infrastructure et utilisateurs de données.

Dans son ouvrage *Seeing like a state*⁴, James Scott démontre le lien entre la construction des Etats modernes et la standardisation. Mise en œuvre au travers de divers instruments (statistique, cadastre, etc.) par l'appareil bureaucratique, la standardisation réduit les diversités culturelles et simplifie l'hétérogénéité des savoirs locaux. Elle crée ainsi une société homogène, davantage « lisible » et gouvernable par l'Etat central. En centralisant et standardisant les données de producteurs divers, la plateforme participe, selon ce même schéma, à ce processus de construction de nouvelles autorités politiques. Elle réduit

¹ Pierre Collin et Nicolas Colin, *Mission d'expertise sur la fiscalité de l'économie numérique*, Paris, Ministère de l'économie et des finances, 2013.

² Nicholas Carr, *Google in the middle*, <http://www.routhtype.com/?p=1249> , 10 avril 2009, (consulté le 15 janvier 2015).

³ Kevin Boudreau et Andrei Hagiu, « Platform Rules: Multi-Sided Platforms as Regulators » dans Annabelle Gawer (ed.), *Platforms, Markets and Innovation*, Northampton, Edward Elgar Publishing, 2011, .

⁴ James Scott, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, New Haven, Yale University Press, 1998, 445 p.

l'hétérogénéité des données et redistribue le pouvoir des producteurs de données, en altérant leur autonomie, vers l'acteur possédant la plateforme. La plateforme permettrait ainsi la domination centralisée d'un acteur par la réduction de la diversité et de l'hétérogénéité des points de vue sur l'espace urbain.

III. Articulation

L'articulation est la troisième modalité de réutilisation des données ouvertes. Elle se caractérise par la mise en relation de données hétérogènes par l'intermédiaire d'un attribut commun. L'articulation produit une nouvelle représentation de l'espace urbain par l'association de données qui conservent toute leur hétérogénéité.

Ce type de réutilisation des données ouvertes s'inscrit dans le mode de quantification des *big data*. Le terme de *big data* est employé pour désigner des formes de quantification très variées¹. En m'appuyant sur la définition classique des 3V² (le volume, la vélocité et la variété) qui le caractérise, je le définis ici comme l'usage de bases de données massives et exhaustives, actualisées en temps quasi réel et composées de données hétérogènes³. Le volume souligne les capacités de collecte et de traitement d'un nombre potentiellement

¹ Rob Kitchin et Gavin McArdle, « The diverse nature of big data », *Big Data & Society*, 2016, vol. 3, n° 1, p. 1-10.

² Ces trois caractéristiques ont été mises en avant par une note du cabinet Gartner en 2001. Elles sont aujourd'hui largement utilisées pour décrire ce mode de quantification. Certains auteurs ajoutent néanmoins des dimensions supplémentaires. Outre les trois V, Rob Kitchin propose par exemple de définir le Big data par l'exhaustivité, la granularité, la relationalité et la flexibilité. Doug Laney, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Stamford, META Group, 2001 ; R. Kitchin, *The data revolution: Big data, open data, data infrastructures and their consequences*, *op. cit.*

³ Le *big data* entraînerait également une « rupture épistémologique ». Dans un article volontairement polémique publié en 2008, Chris Anderson a affirmé que l'exhaustivité des données rendrait suffisante la simple constatation de corrélations plutôt que la recherche de causalités. Ce point fait l'objet de nombreux débats sur lesquels je ne reviendrai pas ici. dannah boyd et Kate Crawford ont notamment publié une réponse critique à l'article de Chris Anderson.

Chris Anderson, « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », *Wired*, 23 juin 2008 ; dannah boyd et Kate Crawford, « Critical questions for big data », *Information, Communication and Society*, 2012, vol. 15, n° 5, p. 662-679.

illimité de données¹. Cette exhaustivité² marque une rupture avec les méthodes de quantification précédentes, puisque contrairement aux statistiques, il ne serait plus nécessaire de recourir aux techniques d'échantillonnage. « *Ce n'est pas le nombre absolu de points de données en soi, ni la taille de l'ensemble de données, qui font que l'on classe ces exemples dans la catégorie des big data : cela tient plutôt à l'utilisation de la plus grande quantité possible de données sur la totalité, et non celle d'un raccourci d'un échantillonnage aléatoire*³ ». La vélocité représente la fréquence de production et d'usage des données. Les données sont générées et analysées en flux continu, actualisées en quasi temps réel. Cette focalisation sur le temps court souligne l'inscription principale du *big data* dans un cadre d'action, au contraire de la temporalité plus longue de bases de données consolidées qui offrent davantage de prises à l'analyse et à la compréhension. Avec le *big data*, la réactivité se substitue à la réflexivité. La variété souligne l'hétérogénéité des données traitées. Les données ne sont plus nécessairement des données relationnelles traditionnelles : il peut s'agir de traces, de données semi ou non structurées, enregistrées sous des formats divers et relevant de secteurs variés⁴. Le point important à souligner, sous-entendu dans cette caractéristique de la variété, est la dissociation de la donnée et de son usage : soit la donnée est assignée à d'autres fins (usage secondaire), soit la donnée n'est pas produite pour une finalité particulière (traces numériques).

A. Associer des jeux de données hétérogènes pour produire une nouvelle représentation de l'espace urbain

Pour illustrer ce mode de réutilisation des données par articulation, j'analyse principalement trois cas, portant sur des secteurs d'action publique distincts, mais qui utilisent des données volumineuses, d'une très grande diversité et, pour deux de ces cas, actualisées en

¹ Ces capacités sont le résultat de plusieurs évolutions techniques : la baisse des coûts de stockage, l'augmentation des capacités de calcul (calcul distribué, etc.), la miniaturisation des capteurs, l'émergence des bases de données non relationnelles, etc.

² La référence à l'exhaustivité doit cependant être entendue avec prudence, puisque, comme le rappelle Dominique Boullier, il n'existe aucune « totalité » de référence à laquelle comparer ces jeux de données. D. Boullier, « Vie et mort des sciences sociales avec le big data », art cit.

³ V. Mayer-Schönberger et K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, op. cit., p. 39.

⁴ Je m'intéresse ici particulièrement à la caractéristique de la variété, qui est cruciale pour comprendre le travail d'articulation nécessaire à la mise en place de projets de *big data*. Si le volume et la vélocité transforment également les modes de représentation, ils ne seront traités que secondairement dans mon analyse.

continu pour gouverner la ville. Le premier cas est celui d'*Optimod*, un système de transport intelligent (ITS) mis en place par la communauté urbaine de Lyon, qui rassemble, articule et analyse les données de l'ensemble des modes de transport de l'agglomération pour fournir des services d'information multimodaux aux usagers. Le deuxième cas est celui du *Mayor Office of Data Analytics* (MODA), un service de la ville de New York qui se décrit comme un "civic intelligence center", centralisant, agrégeant et analysant des données provenant de l'ensemble des services de la ville "to prioritize risk more strategically, deliver services more efficiently, enforce laws more effectively and increase transparency"¹. Le troisième cas est un projet de modélisation énergétique des bâtiments du quartier de la Part-Dieu réalisé par la direction de l'urbanisme du Grand Lyon. Il vise à construire une modélisation de la consommation énergétique de chaque bâtiment du quartier pour simuler des scénarios à l'horizon 2030 à partir des consommations actuelles. Cette modélisation s'est appuyée sur des jeux de données hétérogènes provenant du Grand Lyon, de l'agence d'urbanisme de Lyon, d'ERDF et de GRDF. Ce qui rassemble ces cas est leur usage de données de « seconde main » (*second-hand data*) qu'ils articulent pour produire une nouvelle représentation de l'espace urbain.

A partir de données provenant de plusieurs administrations municipales, le MODA souhaite effectuer une analyse prédictive des immeubles présentant des risques de conversion illégale. L'idée est que l'association des points de vue partiels de chacune des administrations, de ces oligoptiques², fasse émerger une représentation plus riche de la ville³. Chaque administration produit des données selon son propre système de catégorisation. En fonction de leur secteur d'intervention, les administrations représentent le *monde* de différentes manières. Un même élément peut donc apparaître dans plusieurs bases de données mais de manière différente. Les administrations ne regardent pas la même chose lorsqu'elles encodent un même élément comme le souligne Michael Flowers, *Chief Data analytics*, de la ville de New York.

The City of New York knows so much about persons, places, and businesses through its regulatory activity. The example I like to give is the coffee shop that you go into. If you walk into any coffee

¹ Source : <http://www.nyc.gov/html/analytics/html/home/home.shtml> (Consulté le 28 octobre 2014)

² B. Latour et Hermant, *Paris, ville invisible*, op. cit.

³ La prétention n'est toutefois pas de construire un panoptique, les représentations issues de ces associations resteront en effet partielles.

shop, usually, you are going to see six or seven stickers on the door. The lion's share of those stickers are going to be generated by a municipality, whether it's the Department of Health or if they serve food; the Fire Department if there's people publically congregating; the Department of Buildings which is concerned about architecture and structural integrity, et cetera, et cetera, et cetera¹.

Le projet de Centrale de mobilité vise également à mettre en relation des données de différentes organisations. Il s'agit dans ce cas d'articuler les données de l'ensemble des modes de transport de l'agglomération lyonnaise pour produire une représentation métropolitaine de la mobilité. Ces données sont également très différentes selon les modes de transport. Les réseaux routiers et de transport en commun sont représentés de deux manières incommensurables (cf. Chapitre 1), rendant impossible l'association de ces deux bases de données dans une base commune homogénéisée.

Ces cas traitent ainsi de données héritées, porteuses de l'empreinte institutionnelle de leur producteur. Ces projets cherchent ainsi à assigner aux données un usage secondaire pour lequel elles n'ont pas été conçues. Les différences de structuration rendent impossible l'homogénéisation des données dans un modèle commun. L'enjeu est de préserver l'héritage de la donnée en interconnectant les jeux de données sans modifier les façons dont ils sont produits. Un travail d'articulation des jeux de données autour d'un dénominateur commun est alors nécessaire.

B. Le travail d'articulation

La mise en relation des données hétérogènes, qui est au cœur du *big data*, nécessite un « travail d'articulation ». Cette notion a été introduite par Anselm Strauss² pour désigner les opérations permettant la coordination entre les différentes tâches dans le parcours de soin d'un patient. Strauss met en avant le caractère problématique des moments d'articulation où se rencontrent des cultures et des points de vue différents et qui contribuent à transformer les caractéristiques des entités mobilisées dans ces situations.

¹ Source : Michael Flowers, *Chief Data Analytics* de la municipalité de New York, cité par Alex Howard, *Predictive data analysis is saving lives and taxpayer dollars*, <http://radar.oreilly.com/2012/06/predictive-data-analytics-big-data-nyc.html>, 26 juin 2012, (consulté le 23 octobre 2015).

² Anselm Strauss, « Work and the division of labor », *The Sociological Quarterly*, 1985, vol. 26, n° 1, p. 1-19 ; Anselm Strauss, « The articulation of project work: an organizational process », *The Sociological Quarterly*, 1988, vol. 29, n° 2, p. 163-178.

De manière similaire, la mise en relation de jeux de données hétérogènes implique d'articuler des points de vue multiples pour produire une nouvelle représentation de l'espace urbain. Une caractéristique centrale du travail d'articulation est son inscription dans un processus de travail : il est toujours entrepris en vue d'une finalité particulière. Une seconde particularité du travail d'articulation est qu'il repose sur une série de négociations¹. Il soulève tâtonnements et incertitudes puisqu'il nécessite de traiter simultanément les problématiques d'héritage et de voisinage des données². Cette opération est le fruit d'un long processus d'identification et/ou d'insertion d'un dénominateur commun entre des données provenant d'acteurs aux cultures diverses. Le travail d'articulation nécessite de prendre le temps d'analyser chaque jeu de données, en prenant en compte la culture métier qui a déterminé sa structuration, afin d'associer, aux travers des données, les points de vue de différents acteurs sur l'espace urbain.

1) Les « opérateurs d'articulation » : la donnée-pivot

L'articulation des données hétérogènes s'effectue par le biais d'« opérateurs d'articulation » tels que les « données pivot ». La « donnée pivot » est un identifiant unique servant à relier des bases de données. Il s'agit de conventions partagées au delà des organisations et des cultures métiers. Les données pivots sont de natures très diverses selon les secteurs d'activités et les entités représentées. On peut citer pêle-mêle les adresses, le cadastre, les noms de lieux, les limites communales, les codes postaux, les IRIS, les registres SIREN, SIRET ou Chorus, les répertoires ISBN, la date et l'heure, etc.

Ces informations, qui font « référence », sont des conventions instaurées majoritairement par des organismes publics qui en assurent la définition, la production et la maintenance. Ils sont garants de leur robustesse, leur stabilité et leur pérennité. L'INSEE a par exemple la charge de la production des registres des entreprises SIREN et SIRET. En construisant un espace d'équivalence, ces normes identificatrices sont indispensables à

¹ A. Strauss, « The articulation of project work: an organizational process », art cit.

² Les notions de voisinage et d'héritage sont proposées par Dominique Boullier pour désigner respectivement les degrés d'immunité et d'irréversibilité des entités. Dominique Boullier, *La ville événement*, Paris, Presses Universitaires de France, 2010, 112 p.

l'accord et la coordination d'acteurs pluriels par le partage d'un même répertoire de langage et d'action. Ces conventions stabilisent la définition d'une entité et de ses qualités.

Cependant, ces références ne font pas toujours conventions. Les données mises à disposition sont celles produites et utilisées quotidiennement au sein d'institution, où l'enjeu de la coordination entre cultures hétérogènes n'est pas aussi prégnant. Ils n'éprouvent alors pas le besoin de s'appuyer sur une référence commune à différentes organisations. L'insertion des données pivots dans ces jeux de données n'est ainsi pas assurée.

Certaines informations, a priori similaires, peuvent s'exprimer de manières très hétérogènes. Par exemple, l'adresse d'un immeuble ne sera pas la même dans les jeux de données de La Poste, d'ERDF, ou de la communauté urbaine de Lyon. Il est alors nécessaire d'harmoniser ces jeux de données en insérant une convention commune (l'adresse) pour articuler ces données comme en témoigne la responsable du projet de modélisation énergétique des bâtiments du quartier Part-Dieu.

ERDF et GRDF, ils ont des adresses, enfin, ils ont des bases bâtiments, qui ne correspondent pas aux adresses postales. Nous, on a des recensements, le bâti, il est rattaché à une adresse postale, ce qui n'est pas le cas chez les énergéticiens, parce que le bâti, il est rattaché au point de raccordement, et le point de raccordement, il n'est pas forcément sur la rue. Si un bâtiment est entre deux, voire entre quatre rues, le point de raccordement, il peut être là, ou il peut être là, ou il peut être là. Donc, on s'est aperçu qu'il fallait déjà harmoniser tous ces fichiers, ce qui est un énorme travail pour ne pas se tromper dans les consommations que l'on affecte au bâtiment. Donc, pour chaque bâtiment, on a du trouver une identification du bâti, que tout le monde puisse utiliser. C'est vrai, parce que pour nous, si on dit le 4 rue Garibaldi, pour nous, ça correspond à un immeuble, pour ERDF, ça peut correspondre à rien du tout parce que ça se trouve leur point de livraison, il n'est pas sur la rue Garibaldi, il est à côté¹.

L'harmonisation de l'attribut « adresse » de ces données s'avère indispensable à leur articulation. Toutefois, il est fréquent que les référentiels soient multiples, et qu'aucun ne fasse réellement convention pour l'ensemble des acteurs. Ce réutilisateur des données ouvertes par différentes collectivités pointe ainsi l'hétérogénéité des référentiels de voirie. Tout l'enjeu est alors de parvenir à traduire ces différentes manières de localiser un même élément dans un référentiel commun.

¹ Source : Entretien avec Marine, chargé de mission, Grand Lyon (19 janvier 2015)

Je dirai que la principale difficulté est l'absence de référentiel commun. Chaque collectivité a son propre référentiel de voirie, et il diffère de celui des bases internationales comme celles de Google, Mappy ou OpenStreetMap. L'enjeu est vraiment celui de la mise en relation de ces différents référentiels. Nous on essaye de pousser les collectivités à utiliser le modèle de référentiel d'OSM, qui est un référentiel libre, ouvert, et international. Ou a minima, de mettre en place des passerelles entre les référentiels¹.

Le choix d'un référentiel commun et l'intégration de celui-ci dans les différents jeux de données nécessitent un important travail préalable à l'articulation des données. Chaque acteur a ses propres méthodes pour représenter l'espace urbain, même pour des informations supposément partagées par tous comme la localisation. Le MODA a développé un outil permettant d'associer les différentes manières de localiser une entité à un identifiant BBL/BIN (*Borough Block Lot / Building identification number*). Le BBL est un système de numérotation unique des parcelles cadastrales de la ville de New York, tandis que le BIN est un numéro unique assigné par le *Department of City Planning* à chaque bâtiment. Ces deux identifiants sont utilisés par de nombreuses administrations municipales ; ils sont une convention, qui n'est néanmoins, encore une fois, pas partagée par tous.

Each agency has its own ontology of terms and data that have all been created through reasonable, rational evolution of service, but which sometimes make it nearly impossible to connect that data. One department may use a GIS identifier for the location of the downed tree, whereas another may refer to it by its cross streets. For us, we found that BBL/BIN (borough block lot/building identification number), along with a specialty geocoding software program one of our analysts wrote, was the Rosetta Stone to connecting the city's operational intelligence. For most city agencies, BBL and BIN are the standard way of identifying a location; however, they're not used by all agencies, nor are they universally appropriate. However, we can take whatever geo data we have (an address, an intersection, etc.) and geocode it to the nearest BIN/BBL. By focusing on the common denominator, which is structures in specific locations in this case, we're able to tie together datasets that have previously never been linked².

En attribuant à chaque information géographique un numéro BBL/BIN, le service de *data analytics* de la ville de New York construit un dénominateur commun entre les jeux de données. Cette opération, loin d'être aisée, rend possible l'assemblage de jeux de données hétérogènes, et ainsi, la production d'une nouvelle représentation de la ville à partir de différentes perspectives sur l'espace urbain.

¹ Source : Entretien téléphonique avec Paul, ingénieur, Géovélo (7 juillet 2014)

² Michael Flowers, « Beyond Open Data: The Data-Driven City » dans Brett Goldstein et Lauren Dyson (eds.), *Beyond Transparency*, San Francisco, Code for America Press, 2013, p. 185-199.

De manière similaire, dans le cadre du projet Optimod, un référentiel commun a été constitué permettant de mettre en relation les différents réseaux de transport de l'agglomération lyonnaise. L'enjeu était d'harmoniser les références géographiques de chaque jeu de données provenant des opérateurs pour les relier. Concrètement, ce référentiel géographique prend la forme d'une base de données décrivant l'ensemble du réseau de voirie de l'agglomération sur lequel les différentes bases de données viennent s'articuler. Il établit tout d'abord les correspondances entre les transports collectifs pour un même arrêt desservi par différents opérateurs. Ce référentiel d'offre théorique de transport en commun (arrêts, tronçons, lignes et données horaires) est ensuite calé par rapport à un référentiel routier de voirie. Ces opérations de *map-matching* sont fastidieuses et doivent être renouvelées à chaque mise à jour des données initiales.

Chacune des données est conçue de manière autonome, et non pour communiquer entre elles. Donc, pour traiter l'hétérogénéité des données, on met en place une carte multimodale. L'enjeu est d'avoir un référentiel commun entre les données routières et les données de transport en commun. [...] Il faut rassembler ces données sur une base navigable, c'est-à-dire composée de plusieurs segments, des tronçons de voies, qui sont reliés entre eux. [...] [Cette] base de données comprend l'ensemble des informations qui sont nécessaires pour faire de la navigation et notamment les sens de circulation. Après, pour chaque couche de donnée, on a besoin de la "matcher", c'est-à-dire de faire correspondre l'information de la donnée à un des tronçons de cette base. On « cartographie » les données pour les associer à un tronçon. Et, pour ça on a mis au point des algorithmes permettant de faire la correspondance entre les référentiels géographiques utilisés sur chaque couche. On en a un pour le PC Criter, un autre pour les vélos, également pour les parkings. Maintenant, on a un système semi-automatique. On a commencé à faire ça à la main pour chacune des informations, mais c'était très long. Puis, on a fait ça de manière complètement automatique, mais il y avait pas mal d'erreurs. Donc, maintenant, on fait ça automatiquement, mais on vérifie systématiquement manuellement¹.

Dans ces deux cas, les données sont mises en relation, mais restent complètement indépendantes. Elles ne sont pas transformées ou homogénéisées, elles conservent leur spécificité. Chaque institution conserve la maîtrise de la production de sa base de données et reste responsable de sa mise à jour et de sa fiabilité. Ce système de « data bridge » au travers de « gateways² » permet d'interconnecter des systèmes informationnels initialement autonomes et incompatibles. Ce travail de mise en relation est spécifique à chacune des bases de données initiales pour lesquelles un algorithme de correspondance a été développé. Il ne

¹ Source : Entretien téléphonique avec Alexis, Responsable du référentiel, Cityway (9 juillet 2014).

² Paul A. David et Julie Ann Bunn, « The economics of gateway technologies and network evolution: Lessons from electricity supply history », *Information Economics and Policy*, 1988, vol. 3, n° 2, p. 165-202.

peut être cependant totalement automatisé et un travail manuel de vérification est toujours nécessaire.

Une autre difficulté du travail d'articulation réside dans la diversité des échelles de représentation des données. Pour déterminer un dénominateur commun, il est alors nécessaire de négocier avec les jeux de données, soit en agrégeant les informations à une échelle supérieure, soit en les désagrégeant à une échelle inférieure par des calculs de probabilité. Dans cet exemple, cet acteur bricole la donnée afin de désagréger l'îlot en bâtiments. La fiabilité des estimations obtenues varie selon les méthodes et l'usage que l'on souhaite faire de ces données articulées.

Dans une certaine mesure, quand on dispose de données à une certaine échelle, on peut les désagréger au bâtiment, ou à une échelle inférieure en fonction des thématiques : en fonction du nombre d'habitants, en fonction des bâtiments, en fonction de différents paramètres, où là on sait que le degré d'erreur, il est considérable en fonction de la méthode de désagrégation que l'on a adoptée¹.

Le travail d'articulation repose ainsi sur une série de négociations avec les jeux de données pour parvenir à faire tenir ensemble des données dont les échelles de représentations divergent. Il faut « négocier » avec les jeux de données pour faire coïncider différents niveaux. Ces opérations laborieuses d'articulation de données hétérogènes soulignent les « data frictions² » inhérentes dès lors que des données sont assignées à un autre usage que celui pour lequel elles ont été conçues.

Ce « travail démentiel³ » de mise en relation de données hétérogènes peut toutefois limiter les réutilisations de données comme le souligne la responsable du projet de modélisation énergétique du quartier Part-Dieu.

On voulait voir si cette méthode était fiable pour pouvoir ensuite la dupliquer sur l'ensemble de l'agglomération. Parce que le travail que l'on a fait sur le quartier de la Part Dieu, et qui a été fait aussi avec des données réelles, récupérées chez ERDF et GRDF et sur le réseau de chaleur et de froid, c'est un travail démentiel ! Qui a demandé des heures et des heures ! [...] Pour avoir des fichiers propres, nets, c'est un énorme travail, donc en fait, les outils de modélisation, c'est une

¹ Source : Entretien avec Manon, ingénieur SIG, Forcity (5 mars 2015)

² P. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.*, op. cit.

³ Source : Entretien avec Marine, chargé de mission, Grand Lyon (19 janvier 2015)

vraie avancée pour arriver à faire des scénarios, mais pour arriver à avoir des bases de données d'entrée qui soient exploitables, c'est juste un travail de fou¹.

Face à l'ampleur de la tâche d'articulation, le projet de modélisation énergétique mené à l'échelle du quartier de la Part Dieu ne sera par exemple pas déployé sur l'ensemble de l'agglomération. « *On imagine pas que ce travail que l'on a fait sur Part Dieu, il puisse être reproductible sur l'ensemble de l'agglomération, parce que c'est trop de travail*². » Ainsi, le travail d'articulation représente un investissement important, souvent peu visible, mais qui peut contraindre les réutilisations des données ouvertes.

Encadré 14 : L'essor des data scientists, de nouveaux acteurs administratifs ?

Ces opérations d'articulation, qui nécessitent des compétences spécifiques, font émerger de nouveaux acteurs participant au travail administratif. La ville de New York a créé de nouveaux postes pour nettoyer, lier et analyser les données. Ces *data scientists* sont principalement des jeunes professionnels (avec moins de cinq ans d'expérience de travail), ayant une formation académique en mathématiques ou en économie³. Le Grand Lyon ne dispose pas aujourd'hui de *data scientists*. Un travail similaire d'articulation et d'analyse de données est effectué occasionnellement par les services géomatiques des directions. Dans le cadre du projet Optimod, la communauté urbaine a eu recours à l'entreprise Cityway pour constituer le référentiel.

Pour faciliter ces opérations d'articulation de jeux de données hétérogènes, les producteurs peuvent insérer des « opérateurs d'articulation » au sein de leurs jeux de données. L'ajout de cette donnée-pivot permet de dépasser le caractère localisé de la donnée et de l'inscrire dans un espace d'équivalence commun aux données d'autres producteurs. A l'échelle des collectivités territoriales, cela s'effectue principalement par l'ajout d'identifiants géographiques qui servent d'opérateurs d'articulation (code INSEE ou IRIS, numéro unique d'un bâtiment, etc.), mais cela peut être étendu aux entreprises (code SIREN/SIRET) voire

¹ Source : Entretien avec Marine, chargé de mission, Grand Lyon (19 janvier 2015)

² Source : Entretien avec Marine, chargé de mission, Grand Lyon (19 janvier 2015)

³ Pour une analyse critique de l'émergence des ces nouveaux professionnels de l'économie de la connaissance, voir Robert Gehl, « Sharing, knowledge management and big data: A partial genealogy of the data scientist », *European Journal of Cultural Studies*, 2015, vol. 18, n° 4-5, p. 413-428.

aux individus (numéro de sécurité sociale). L'universalité de cet identifiant déterminera l'étendue du réseau auquel pourra être reliée la donnée. Par exemple, pour élargir le périmètre d'usage de sa donnée, un producteur du Grand Lyon envisage d'ajouter un attribut supplémentaire dans sa donnée : un identifiant unique correspondant à une convention nationale stabilisée.

La localisation, c'est bien, mais il n'y a pas d'intelligence métier dessus. Tu ne peux pas facilement dire, il y a tant de classes dans cette école ou il y a tant de camions de pompiers dans cette caserne. Si je veux que des réutilisateurs s'approprient cette information et l'enrichissent, mon idée est d'ajouter sur ces objets, en plus de leur localisation et de leur nom, un champ qui serait un identifiant externe. Typiquement sur les écoles, il y a un numéro national qui doit être attribué par le ministère de l'éducation nationale et chaque école, je suis persuadé qu'elle a un identifiant unique. Ce que je voudrais que l'on arrive à faire dans le service l'année prochaine, c'est pour une partie de ces sous-thèmes, récupérer un identifiant intelligent, l'ajouter dans notre base et le diffuser derrière. En gros, ça veut dire que si demain, j'ai les crèches, avec un numéro que la préfecture donne, n'importe qui va pouvoir joindre, sur cet identifiant de l'information qu'il aura récupérée auprès de la préfecture, ou je ne sais pas qui, avec le nombre de places de la crèche, et des taux de fréquentation, à eux d'inventer quelque chose. Mais en tout cas, ils pourront, s'appuyer sur cette base de données là, pour y adjoindre des éléments que moi je n'ai pas, et qu'ils pourront avoir par d'autres canaux : un ministère, une préfecture, ou n'importe qui d'ailleurs¹.

Cet ajout d'un identifiant unique, qui vise à faciliter le travail d'articulation, n'est pas anodin sur les modalités de production de la donnée². *« Là, quelque part, je ne peux pas dire le contraire, c'est directement dans une optique de réutilisabilité que je fais ça, c'est pas dans une optique d'usage interne. Par contre, ça va m'obliger à les saisir. A trouver l'info, ce n'est pas forcément simple. Et à la saisir. Et on a comme ça, 5000 ou 10 000 objets à reprendre³ »*. Un investissement, coûteux, est nécessaire pour ajouter cette information à la donnée afin de faciliter son association à d'autres jeux de données. En ajoutant ce nouvel attribut, le producteur participe au travail d'articulation et favorise la réutilisabilité de sa donnée.

2) Un travail d'articulation spécifique à chaque enjeu traité

Si ce travail d'articulation est laborieux, il doit en outre être repris pour chaque usage qui est fait de ces données associées. En fonction du problème à traiter, l'articulation des

¹ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

² Une évolution de l'outil de saisie de la donnée est nécessaire pour contraindre les opérateurs à renseigner l'identifiant unique.

³ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

données est recomposée : le dénominateur commun choisi diffère. Par exemple, si le numéro BBL/BIN a été le premier « opérateur d'articulation » utilisée par le MODA, en fonction des finalités poursuivies, de nouvelles manières d'associer les données doivent être identifiées.

One of the things we have learned is that the way you put data together is specific for what you want to use it for. And so, the effort to bring everything on a common location platform is very useful. But, we have to change very frequently around some of the specific needs that we have. If we are doing something that is associated to the building, having a building identifier for that, and all of the information type of that, is totally fine. If we are doing something, let's say emergency response, we need to know the specific address, and there are a lots of buildings that are around the corner, identifying the building as a whole is no longer sufficient, you need to identify a specific entrance. And so, after having built a lot of these connections, we realize that's not the end. That's a good starting point, but it needs customization understanding the information in order to bring it to bear on a project¹.

De manière similaire, dans le cas de la modélisation des consommations énergétiques précédemment présenté, l'expérimentation a permis de mettre en avant le fait que le bâtiment n'est pas l'échelle la plus adéquate pour mettre en œuvre le schéma directeur énergie de l'agglomération. Si l'identifiant IRIS semble être le dénominateur commun pertinent pour établir un schéma directeur à l'échelle de l'agglomération, il ne présente que peu d'intérêt lorsque l'on souhaite orienter les politiques sur la rénovation des bâtiments et cibler les personnes en précarité énergétique. Il est alors nécessaire d'articuler les données selon un dénominateur plus fin (le bâtiment, le foyer, etc.). L'articulation des données est à géométrie variable selon les phénomènes à traiter. La question de l'échelle d'articulation est toujours associée à un usage, elle est recomposée pour chaque enjeu que l'on souhaite gouverner.

*

* *

En gérant les conséquences de la nature distribuée du travail² (Star et Strauss, 1999, p. 10), le travail d'articulation est essentiel pour faire tenir ensemble des données produites au sein de « silos » métiers hétérogènes pour répondre à des finalités diverses. Comme le mettent en avant Gerson et Star, localisé et temporaire, il s'inscrit toujours dans une perspective d'action.

¹ Source : Entretien avec Nick, Chief Data Analytics, New York City (30 octobre 2014)

² Susan Leigh Star et Anselm Strauss, « Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work », *Computer Supported Cooperative Work*, 1999, vol. 8, n° 1-2, p. 10.

Reconciling incommensurate assumptions and procedures in the absence of enforceable standards is the essence of articulation. Articulation consists of all the tasks involved in assembling, scheduling, monitoring, and coordinating all of the steps necessary to complete a production task. This means carrying through a course of action despite local contingencies, unanticipated glitches, incommensurate opinions and beliefs, or inadequate knowledge of local circumstances. [...] Every real world system thus requires articulation to deal with the unanticipated contingencies that arise. Articulation resolves these inconsistencies by packaging a compromise that 'gets the job done,' that is closes the system locally and temporarily so that work can go on¹.

A l'instar de ce que pointent Gerson et Star, en l'absence de standardisation, l'articulation des données ouvertes est entravée par plusieurs sources de perturbations : absence d'attribut commun, échelles discordantes, classifications divergentes, etc. Tous ces aléas obligent les acteurs à négocier avec les bases de données pour parvenir à les rendre commensurables. Ce travail d'arrière-plan, préalable nécessaire à toute analyse basée sur le *big data*, est pourtant largement sous-estimé. Peu formalisées, ces opérations sont ensuite « invisibilisées² » afin de faire tenir les représentations qui émergent de ces associations de données. Comme le souligne Florence Millerand, le travail d'articulation se caractérise par des « *activités de bricolage, de traduction et d'effacement*³ » : bricolage avec les bases de données, traduction des points de vue de mondes sociaux hétérogènes et effacement de ce travail pour assurer la solidité de la nouvelle représentation. Ces caractéristiques sont essentielles pour parvenir au « *compromise that 'gets the job done'* » pour reprendre les termes de Gerson et Star. L'inscription dans un cadre d'action est une propriété essentielle du travail d'articulation, qui particularise le *big data* comme politique de quantification toujours orientée vers l'action.

¹ Elihu M. Gerson et Susan Leigh Star, « Analyzing due process in the workplace », *ACM Transactions on Office Information Systems*, 1986, vol. 4, n° 3, p. 266.

² Cette mise en invisibilité est similaire à celle du travail effectué par les gestionnaires de bases de données, comme le soulignent Eric Dagiral et Ashveen Peerbaye.

Eric Dagiral et Ashveen Peerbaye, « Les mains dans les bases de données. Connaître et faire reconnaître le travail invisible », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 191-216.

³ Florence Millerand, « La science en réseau. Les gestionnaires d'information « invisibles » dans la production d'une base de données scientifiques », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 163-190.

C. Nouvelle représentation de l'espace urbain, nouvel objet de gouvernement

A l'issue de ces opérations d'articulation, une nouvelle représentation de l'espace urbain émerge. En associant différents points de vue sur l'espace urbain, le travail d'articulation transforme les attributs des entités représentées. Il permet de déployer et de représenter le réseau spécifique qui caractérise chacune des entités articulées. Par exemple, dans le cas du MODA, les bases de données sur les bâtiments détenues par les différents services municipaux sont articulées. Il est alors maintenant possible d'obtenir pour chaque bâtiment toutes les informations le concernant (permis de construire, titre de propriété, plaintes, etc.).

Now, for the first time, I can search, let's say 253 Broadway, and in one place, I can find the permits on it, I can find the water use, I can find the violations, I can find the owner, I can find the building history, I can find all these information for the building¹.

Chaque immeuble est dorénavant le centre d'un réseau composé des jeux de données municipaux le concernant. A partir de l'articulation de multiples points de vue, il est alors possible par des dispositifs de calculs de singulariser chacune de ces entités qui deviennent autant de cibles de gouvernement. En partant du postulat qu'il existe un caractère régulier et prévisible des pratiques², que le futur est une reproduction du passé, les *data analysts* du MODA construisent des *patterns* qui leur permettent de distinguer les immeubles à risque et de cibler les inspections municipales.

Pour étudier plus précisément le changement des modes de gouvernement induit par l'articulation de bases de données hétérogènes, j'ai choisi d'étudier plus précisément le cas du projet Optimod. L'articulation des données de l'ensemble des transporteurs recompose la représentation de l'espace des transports urbain. En permettant la singularisation des représentations, il souligne le passage du gouvernement des transports au gouvernement de la mobilité métropolitaine.

¹ Source : Entretien avec Nick, Chief Data Analytics, New York City (30 octobre 2014)

² Pour une analyse sociologique des algorithmes, voir Dominique Cardon, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Seuil, 2015, 112 p.

1) Du gouvernement des transports au gouvernement de la mobilité

En associant les données de l'ensemble des opérateurs de transport de l'agglomération lyonnaise, la « Centrale de mobilité » offre une nouvelle représentation de l'espace de la mobilité urbaine. L'objectif de ce projet est d'offrir de nouveaux services d'information aux voyageurs afin de favoriser le report modal de la voiture vers les transports en commun ou les modes doux. Cette nouvelle représentation de l'espace urbain transforme l'objet de la politique publique : on glisse de la régulation des flux des réseaux de transport vers la régulation des individus.

(a) L'utilisateur au centre d'un réseau de données

Pour fournir de nouveaux services d'information, la représentation métropolitaine de la mobilité construite dans la « Centrale de mobilité » est insérée dans des dispositifs de calcul d'itinéraire. Ces artefacts prennent des formes plurielles (sites internet, applications sur téléphone mobile, etc.), mais ils offrent un même service de calcul d'itinéraire multimodal entre une origine et une destination d'un déplacement¹.

Prenons l'exemple de l'application mobile Optymod, première application développée dans le cadre du projet de « Centrale de mobilité » du Grand Lyon. Ce navigateur GPS multimodal utilise les données des différents opérateurs de transport mises en relation par l'intermédiaire d'un référentiel géographique commun. Il ne fournit pas cependant à l'utilisateur l'ensemble des informations présentes dans la Centrale de mobilité, mais uniquement celles qui sont pertinentes pour son déplacement².

¹ Dans sa version la plus simple, le dispositif de calcul d'itinéraire repose sur des algorithmes radicalement différents de ceux utilisés par la ville de New York. En effet, il ne s'agit pas ici de prédire le futur à partir des comportements passés, mais de déterminer un itinéraire entre deux localisations géographiques. Ce n'est que dans ses versions les plus élaborées, où il intègre les traces passées, qu'il peut affiner ce calcul d'itinéraire en précisant l'analyse prédictive des degrés de congestion : par exemple affiner le temps de parcours par mode en fonction du degré prévisionnel d'embouteillage.

² Il ne s'agit pas d'un échantillon de la base de données : la base de données est bien interrogée dans son exhaustivité, mais on ne sélectionne que les informations sur les modes de transport en lien avec la localisation et le déplacement de l'individu. La multiplication de la puissance de calcul informatique rend possible ces opérations dans une temporalité très courte.

Pour cela, il actualise la représentation de l'espace urbain de la mobilité en fonction de plusieurs informations relatives au déplacement de l'individu : son origine et sa destination, sa temporalité, et ses modes de transport privilégiés. Les données relatives aux réseaux de transport sont ainsi croisées avec des informations sur la localisation de l'utilisateur, obtenues soit par renseignement direct par l'utilisateur, soit par géolocalisation par l'intermédiaire de son téléphone mobile¹, mais également avec des informations temporelles (le temps de l'interaction et le temps du déplacement) et des informations identitaires propres à chaque utilisateur (mode de déplacement privilégié, etc.).

La localisation des individus devient l'unique relation spatiale pertinente. La position géographique de l'individu est la référence à partir de laquelle la représentation de la mobilité urbaine est réagencée. Ce réarrangement constant mène à une singularisation des représentations de l'espace urbain qui ne sont plus statiques mais dynamiques². Le navigateur multimodal dispose en outre d'une fonction de suivi de l'utilisateur pendant son trajet, afin de tenir compte de sa localisation en temps réel pour adapter l'itinéraire initialement proposé en fonction des perturbations survenues pendant son déplacement.

On a également sur l'application une actualisation de l'itinéraire en fonction des événements. Il y a un *tracking* de l'utilisateur pour lui indiquer un itinéraire mis à jour s'il est en retard par exemple, et qu'il n'aura plus sa correspondance³.

Les solutions de mobilité présentées à l'utilisateur sont ainsi reconfigurées en permanence en fonction de l'évolution de sa localisation et des événements qui peuvent advenir (Illustration 1). La représentation est devenue dynamique en croisant à la volée des informations temporelles et spatiales qui sont actualisées en temps réel lors du déplacement de l'individu. Cette représentation de l'offre des transports s'inscrit dans une perspective d'action de l'utilisateur. En sélectionnant les informations pertinentes liées à la localisation et

¹ Ou tout autre dispositif permettant une géolocalisation automatique (GPS embarqué des véhicules, navigateurs internet, etc.).

² En agissant sur l'environnement informationnel de l'individu plutôt que sur l'individu et en recomposant en permanence cet environnement en fonction du comportement de l'individu, l'application s'inscrit dans les techniques de gouvernement propre à la « gouvernementalité algorithmique ».

Antoinette Rouvroy et Thomas Berns, « Gouvernementalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? », *Réseaux*, 2013, vol. 1, n° 177, p. 163-196.

³ Source : Entretien téléphonique avec Alexis, Responsable du référentiel, Cityway (9 juillet 2014).

au parcours de l'utilisateur, l'application ne propose pas une représentation globale de la mobilité mais une représentation singularisée de l'offre des transports disponibles dans le cadre de l'action du voyageur. La représentation de l'espace de la mobilité urbaine est recomposée pour correspondre à l'espace d'action de l'individu. Il faut noter que les bases de données ne changent pas lors de ces opérations : elles sont interrogées mais leurs structures restent identiques.

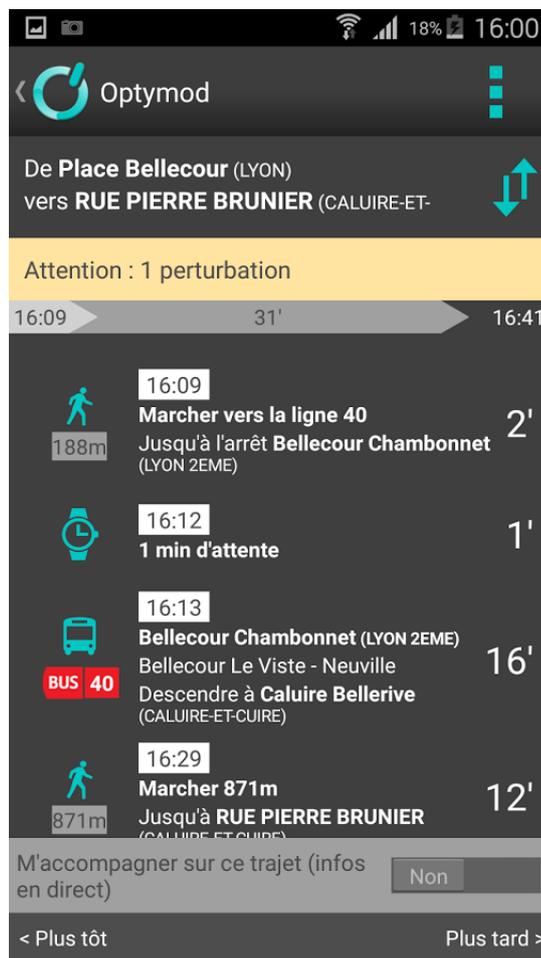


Illustration 29 : Capture d'écran de l'application Optymod

Itinéraire détaillé d'un déplacement calculé à partir de la localisation de son origine et de sa destination. L'itinéraire est actualisé et recomposé en permanence en fonction des événements survenant sur ce trajet (l'alerte sur les perturbations et l'option d'accompagnement en temps réel).

(b) Un usager pluriel plutôt que des flux

Le réagencement dynamique de la représentation de l'espace de la mobilité transforme la cible de gouvernement. La régulation ne vise plus des flux de voyageurs passifs, contraints dans un mode de transport, mais des usagers pluriels aux déplacements particuliers. Ces nouveaux services d'information ne ciblent plus les masses mais des personnes singulières dans toute leur pluralité.

Ce que l'usager recherche effectivement, ce sont des solutions de mobilité intégrée, et on ne va pas le caricaturer en disant, celui qui prend le transport en commun, celui qui prend le vélo, celui qui prend la voiture. Selon les besoins, vous avez des enfants à déposer, vous avez des choses lourdes à transporter, vous ne prenez pas le même moyen de transport. Et donc, il faut pouvoir adresser tout ça¹.

Grâce au croisement avec les informations spatiales et temporelles, l'information de l'offre de transport est actualisée en fonction de la situation dans laquelle se trouve l'utilisateur. Or, celui-ci peut tout à la fois dans une même journée être salarié ou consommateur, seul ou en famille. Le voyageur peut passer d'une posture à l'autre, on refuse tout essentialisme pour lui offrir plusieurs possibilités d'existence. On ne prend plus seulement en compte les modes, mais également les motifs du déplacement. « *On est tous multimodaux, et l'idée c'est de permettre à l'usager de choisir le meilleur mode de déplacement pour le motif qu'il s'apprête à faire*². » L'information mobilité n'est alors plus segmentée par mode de transport, l'usager doit pouvoir « piocher » dans les offres disponibles en fonction de ses besoins. Cela n'est permis que par l'articulation des données et leur actualisation permanente pour proposer une représentation singularisée de l'espace de la mobilité métropolitaine.

Cette singularisation de la représentation marque le passage d'une politique des transports urbains à une politique de la mobilité métropolitaine³. « *Le transport s'adresse à*

¹ Intervention de Jean Coldefy lors de la table-ronde « La mobilité intelligente à l'échelle du bassin de vie », Congrès ATEC ITS France – Les rencontres de la mobilité intelligente, Paris, 28 janvier 2015.

² Entretien avec Guillaume, Responsable mobilité urbaine, Grand Lyon (22 janvier 2014).

³ Pour une histoire de l'évolution des transports à la mobilité, on pourra se reporter à Mathieu Flonneau et Vincent Guigueno (eds.), *De l'histoire des transports à l'histoire de la mobilité ?*, Rennes, Presses Universitaires de Rennes, 2009, 334 p ;

*des flux idéalement homogènes constitués de particules équivalentes. La mobilité, quant à elle, fait d'emblée référence à la personne singulière et à son activité*¹. » Les données étaient utilisées exclusivement pour réguler les flux des réseaux de transport, peuplés d'individus indifférenciés envisagés sous le seul angle de la masse qu'ils composent. La recomposition et le croisement des données permettent dorénavant de porter une attention nouvelle à l'utilisateur et de l'appréhender dans toute sa singularité.

Ce changement de représentation fait évoluer la cible des politiques publiques. Alors qu'auparavant les données étaient des instruments de régulation des réseaux de transport, elles doivent dorénavant permettre à chaque usager, envisagé comme un acteur rationnel, de comparer les modes de transport à sa disposition. La régulation des transports urbains se déplace ainsi des réseaux aux individus².

D. Conclusion : hétérogénéité des données et singularisation des entités représentées

L'articulation est la troisième modalité de réutilisation des données ouvertes. Elle consiste à rendre interopérables des bases de données par l'intermédiaire d'une donnée-pivot : un attribut commun à ces bases de données, qui assure leur mise en relation, défini en fonction de l'enjeu à gouverner. Ce mode de composition des données ouvertes produit une nouvelle représentation de l'espace urbain qui conserve toute la richesse des points de vue de chacune des administrations. Reflet des cultures métiers des producteurs, l'hétérogénéité des données est préservée et même encouragée pour mettre en lumière la pluralité ontologique des entités représentées³.

Plus généralement sur l'émergence du paradigme de la mobilité, voir Mimi Sheller et John Urry, « The New Mobilities Paradigm », *Environment and Planning A*, 2006, vol. 38, n° 2, p. 207-226 ; John Urry, *Mobilities*, Cambridge, Polity Press, 2007, 336 p.

¹ Georges Amar, *Homo mobilis. Le nouvel âge de la mobilité.*, Paris, FYP, 2010.

² Plus généralement, pour l'essayiste Evgeny Morozov, l'idéologie du numérique repose sur la régulation du comportement des individus par le biais de flux d'informations. Evgeny Morozov, *Le mirage numérique : Pour une politique du Big Data*, Paris, Les prairies ordinaires, 2015, 130 p.

³ Chaque entité devient une « monade » : le centre d'un système de relations qui entremêle l'ensemble des points de vue de l'administration sur cette entité. Bruno Latour et al., « 'The whole is always smaller than its parts' – a digital test of Gabriel Tarde's monads », *The British Journal of Sociology*, 2012, vol. 63, n° 4, p. 590-615.

Loin de constituer une simple agrégation d'informations, l'articulation permet la singularisation des entités représentées et transforme les modalités de leur gouvernement. Ces conclusions rejoignent celles d'Evelyn Ruppert dans ses travaux conceptualisant les bases de données des politiques sociales britanniques. « *These databases do not simply add up data about subjects but materialize ontologically different subjects in relation to what they do. [...] Joined-up databases materialize the “individuality” of subjects in intensified, distributed and fluctuating ways*¹. » En rendant visible la diversité de l'entité représentée, l'articulation des bases de données est constitutive de nouvelles façons de la penser et de la gouverner. Contrairement aux modes de quantification précédents, la ville n'est plus observée (et gouvernée) au travers de catégories stabilisées rassemblant un ensemble d'entités aux caractéristiques uniformes². La singularité de chacune de ces entités est dorénavant mise en avant pour adapter le gouvernement à leurs spécificités et aux variations de leur comportement. Par le *big data*, ce n'est plus le « nous » qui est gouverné, mais chaque individu qui devient gouvernable.

IV. Conclusion

L'étude des réutilisations interroge les modalités de coordination entre des mondes sociaux hétérogènes au travers des données ouvertes. Ces coordinations peuvent être intra ou inter organisationnelles : entre producteurs et utilisateurs externes, mais également entre différents services d'une administration qui trouvent au travers de l'open data de nouvelles modalités de coordination dépassant les frontières sectorielles des politiques publiques.

Pour que la coordination réussisse, tout un ensemble d'opérations laborieuses sur les données sont nécessaires : il faut les identifier, les nettoyer, les comprendre, les uniformiser,

¹ E. Ruppert, « The Governmental Topologies of Database Devices », art cit, p. 120.

² Cette focalisation sur l'individu et l'étude de phénomènes conjecturels positionnent le *big data* dans ce que Fabrice Bardet appelle un système de quantification « comptable ». Fabrice Bardet oppose cette quantification au régime probabiliste qui se caractérise par son intérêt pour un groupe social plutôt que pour l'individu, pour la moyenne et la médiane plutôt que pour la singularité.

F. Bardet, *La contre-révolution comptable. Ces chiffres qui (nous) gouvernent*, op. cit.

les restructurer, ajouter des attributs, etc. L'étude et la description de ces opérations permettent de distinguer trois modalités principales de composition avec les données ouvertes : la consolidation, l'homogénéisation et l'articulation. La consolidation consiste à ajouter un ensemble d'opérations à la chaîne de la donnée afin d'assurer sa solidité au sein d'un nouvel environnement. L'homogénéisation est l'usage de données de sources hétérogènes pour produire un nouvel agrégat uniforme. L'articulation est la mise en relation de données diverses au travers d'une donnée-pivot.

Ces trois politiques de données ouvertes ne constituent pas seulement différentes modalités de réutilisation des données. En passant d'un monde social à un autre, la donnée ouverte se dote d'autres qualités. Cette évolution des données modifie les représentations de l'espace urbain qu'elles véhiculent, et, influe, in fine, sur les politiques publiques¹.

- La donnée consolidée est un objet-frontière permettant la coordination d'acteurs pluriels autour d'une représentation commune de la ville. La consolidation offre au producteur de nouvelles capacités de régulation en rassemblant ces acteurs autour de sa donnée.
- La donnée homogénéisée propose une représentation alternative de l'espace urbain par la mise en place d'un nouvel espace d'équivalence qui échappe au producteur initial. Ce dernier perd alors la maîtrise de la représentation de la ville au risque de conflits entre ces points de vue hétérogènes.
- La donnée articulée rassemble une multiplicité de points de vue sur un même objet. Par ce biais, elle rend visible la singularité de chacune de ces entités et offre la possibilité de gouverner spécifiquement chacune d'entre elles.

Ces politiques de réutilisation ne sont pas exclusives². Elles sont déployées sur des périmètres variables, au sein de secteurs d'action multiples et par des acteurs diversifiés. Les

¹ En s'attachant de façon différente, la donnée fait émerger des agencements alternatifs recomposant le gouvernement urbain. Ces agencements mériteraient d'être caractérisés plus finement par une analyse plus précise des modalités de réutilisation des données.

² Une étude plus large des réutilisations des données ouvertes mériterait d'être effectuée pour affiner cette analyse et mettre en avant les distinctions qui apparaissent au sein de ces trois grandes politiques de réutilisation.

réutilisations peuvent concerner un seul secteur d'action publique, ou au contraire, mettre en relation plusieurs secteurs auparavant différenciés. Les échelles sont également multiples : elles peuvent être à l'échelle d'un bâtiment, d'une rue, d'un quartier, d'une commune, d'une métropole, d'une région, d'un pays ou du globe en fonction du degré de granularité et de généralisation choisie. Certains acteurs peuvent s'inscrire dans plusieurs de ces politiques. Google Maps par exemple repose à la fois sur une logique d'homogénéisation par la construction d'une base de données uniforme et une logique d'articulation de données hétérogènes sur la base de ce référentiel géographique.

En l'absence d'identification d'autres utilisateurs des données du Grand Lyon, elle n'a pu être menée dans le cadre de cette thèse.

Chapitre 5

Gouverner : Instaurer une politique de régulation des flux de données

« Votre proposition de thèse tombe à pic, on a un énorme besoin de gouvernance de la donnée¹. »

« Tu vas pouvoir te servir de ça pour montrer à quel point c'est le bordel la gouvernance des données². »

Ces deux citations, extraites de mon journal de terrain, révèlent la préoccupation permanente pour la « gouvernance » de la donnée qui émane des acteurs en charge du projet d'*open data*. La première est tirée de la réunion que j'ai eu avec le directeur des services informatiques du Grand Lyon peu après lui avoir proposé mon projet de thèse CIFRE. La seconde est un échange avec l'équipe projet *open data* qui fait suite à une controverse entre les services de la ville et de la communauté urbaine de Lyon sur un jeu de données similaire produit par chacune des collectivités. La « gouvernance » de la donnée est apparue en toile de fond tout au long de mon observation ethnographique de la politique d'*open data*. Catégorisé comme doctorant en science politique, j'ai été perçu comme une ressource intellectuelle à la mise en place de cette gouvernance. Pourtant, le raccourci « gouvernance / science politique » n'est pas suffisant pour me doter d'une expertise sur le sujet. Comme le relate ce chapitre, mes incertitudes auront été aussi nombreuses que celles de l'ensemble de mes collègues du Grand Lyon. Pour éclaircir ces incertitudes, je souhaite, dans ce chapitre, interroger cette « gouvernance » de la donnée en retraçant sa genèse et en suivant les différentes traductions dont elle a été l'objet au cours des trois années d'observation de la politique d'*open data* du Grand Lyon.

Avant d'aller plus en avant dans la présentation de la « gouvernance » de la donnée, il est nécessaire d'apporter une précision conceptuelle et lexicale. La notion de gouvernance est l'objet de nombreux débats en science politique³. En opposition au « gouvernement », des politistes ont mis en avant ce concept pour souligner le développement des modes d'interactions basés sur la coordination, la négociation et la délibération, plutôt que sur des logiques de domination, de contrainte et d'approches *top-down* qui relèveraient du

¹ Source : Journal de terrain, réunion avec Nicolas, direction des systèmes d'information (16 février 2012)

² Source : Journal de terrain, réunion avec Pauline et François (15 avril 2014)

³ Romain Pasquier, Vincent Simoulin et Julien Weisbein (eds.), *La gouvernance territoriale : Pratiques, discours et théories*, Paris, LGDJ, 2007, 235 p.

gouvernement¹. Par ce schéma d'analyse plus horizontal, les politistes soulignent la pluralité des acteurs engagés, publics ou privés, étatiques ou non-étatiques, dans l'action publique. Outre la polysémie et les usages multiples du terme, les détracteurs du concept de gouvernance pointent la tendance de ces analyses à minorer le pouvoir politique spécifique de l'Etat et les recompositions dont il est l'objet².

L'emploi du terme de gouvernance de la donnée dans ce chapitre ne fait pas référence à ces controverses scientifiques, mais résulte de l'utilisation de ce terme par les acteurs étudiés. A aucun moment, sur mon terrain, le terme de « gouvernement » de la donnée n'a été employé. A contrario, les acteurs font régulièrement référence à la « gouvernance de la donnée ». Par respect pour cet usage, j'emploie ainsi le terme de gouvernance de la donnée pour qualifier indifféremment ce qui pourrait relever d'un gouvernement ou d'une gouvernance telle que l'entend la science politique. L'objet du chapitre n'est pas de déterminer si la politique de la donnée relève du « gouvernement » ou de la « gouvernance », mais d'étudier les modes de coordination mis en place par la Métropole de Lyon pour réguler les données, en étant attentif à l'ensemble des actants qui y sont associés.

Dès lors, plutôt que de partir d'une définition *a priori* de ce que serait le gouvernement ou la gouvernance de la donnée, je souhaite saisir la production de la gouvernance de la donnée « telle qu'elle se fait³ » en m'intéressant à la manière dont les acteurs la définissent et la mettent en œuvre. Ce processus, qui a débuté lors de ma présence au sein du Grand Lyon, est toujours en cours et se caractérise par son incertitude. Je ne vise pas à analyser cette action

¹ P. Le Galès, « Du gouvernement des villes à la gouvernance urbaine », art cit ; Patrick Le Galès, « Gouvernance » dans Laurie Boussaguet, Sophie Jacquot et Pauline Ravinet (eds.), *Dictionnaire des politiques publiques*, Paris, Presses de Sciences Po, 2010, ; Jan Kooiman, *Governing as Governance*, London, Sage Publications, 2003, 256 p.

² Jean-Gustave Padioleau, « L'action publique urbaine moderniste », *Politiques et Management Public*, 1991, vol. 9, n° 3, p. 133-146 ; Bernard Jouve et Christian Lefèvre, « De la gouvernance urbaine au gouvernement des villes ? Permanence ou recomposition des cadres de l'action publique en Europe », *Revue française de science politique*, 1999, vol. 49, n° 6, p. 835-853 ; P. Le Galès, *Le retour des villes européennes*, op. cit. ; Renaud Epstein, « La gouvernance territoriale : une affaire d'État La dimension verticale de la construction de l'action collective dans les territoires. », *L'Année sociologique*, 2015, vol. 65, n° 2, p. 457-482.

³ Michel Callon et Bruno Latour (eds.), *La science telle qu'elle se fait. Anthologie de la sociologie des sciences de langue anglaise*, Paris, La Découverte, 1990, 396 p.

publique déjà constituée et à tirer des conclusions définitives¹, mais à revenir sur les différentes médiations qui ont permis de voir émerger ce gouvernement de la donnée métropolitaine². Je reviendrai tour-à-tour sur les présentations et les textes exprimant le besoin de gouvernance de la donnée, puis sur leur traduction dans des organigrammes et des instruments visant à les opérationnaliser. Ce sont l'ensemble de ces médiations qui permettent la constitution de la donnée comme un enjeu gouvernable et qui légitiment la Métropole de Lyon à s'en emparer.

La construction d'une politique métropolitaine de la donnée repose sur quatre processus que je présenterai successivement avant d'analyser le modèle politique qui émerge de cette politique publique.

- Le premier processus est la formalisation de l'enjeu à traiter : il s'agit de déterminer ce que recouvre la gouvernance de la donnée et d'élaborer des principes généraux d'action.
- Le deuxième est l'institutionnalisation de cette politique, c'est-à-dire son inscription dans l'organisation administrative.
- Le troisième est son instrumentation : la sélection d'outils à utiliser pour la mettre en œuvre.
- Le quatrième est la légitimation, soit rendre acceptable l'autorité de la Métropole de Lyon sur cet objet de gouvernement.

Par ces opérations de formalisation, d'institutionnalisation, d'instrumentation et de légitimation, la direction des services informatiques tente d'ériger la donnée en politique publique autonome dont elle assurerait le pilotage à l'échelle de la Métropole. Pour cela, elle souhaite se positionner en situation d'intermédiaire entre producteurs et utilisateurs de

¹ Avantage et contrainte de la méthode ethnographique, la « gouvernance » de la donnée présentée dans ce chapitre est analysée selon un point de vue situé : celui de la Métropole de Lyon. Mes analyses pourraient être toutes autres si un autre point de vue avait été adopté. En outre, l'analyse est limitée aux prémices de cette politique de la donnée, puisque mes observations s'arrêtent à l'été 2015. Il ne s'agira donc pas de tirer un bilan mais plutôt d'esquisser des pistes d'analyse des politiques de la donnée urbaine.

² L'importance des médiations dans la construction d'une politique publique a notamment été mise en avant par Florian Charvolin dans son étude de la constitution du ministère de l'environnement. Florian Charvolin, *L'invention de l'environnement en France*, Paris, La Découverte, 2003, 133 p.

données. Toutefois, les difficultés rencontrées par cette direction pour réguler la circulation des données révèlent une tension latente entre autonomisation et sectorialisation de cet enjeu de la donnée.

I. Formaliser la gouvernance de la donnée

La première partie de ce chapitre retrace les différentes formalisations de la gouvernance de la donnée apparues au cours de la mise en œuvre de la politique d'*open data* du Grand Lyon. En suivant les acteurs, dans leurs doutes et leurs incertitudes, on parvient à comprendre ce qu'est et ce que n'est pas la gouvernance de la donnée pour la communauté urbaine de Lyon. Formaliser la gouvernance de la donnée, c'est dépasser une idée vague et abstraite de la gouvernance de la donnée en élaborant des principes d'actions qui cadrent la politique publique¹. Préalablement à cela, formaliser la gouvernance de la donnée, c'est identifier les enjeux auxquels devra répondre cette politique publique. En effet, comme le résume Charles W. Anderson : « *Elaborer une politique ne signifie pas uniquement résoudre des problèmes ; mais aussi et d'abord construire des problèmes*² ». Le « problème » de la donnée ne s'impose pas de manière naturelle, évidente et stable. Il est construit par la mobilisation d'acteurs qui sélectionnent les points sur lesquels portera l'action publique, priorisent des enjeux et en écartent d'autres. Le problème public ainsi formalisé est également l'objet de plusieurs traductions : une succession de redéfinitions et de recadrages des formulations initiales qui conduisent à l'association ou la dissociation de certains acteurs à cette politique publique naissante.

¹ J'emprunte ce terme de formalisation à Pierre Lascoumes qui le définit comme « *l'élaboration de principes d'action généraux destinés à cadrer les politiques publiques en dépassant les particularismes nationaux et territoriaux* ». La formalisation est ainsi un processus de sélection et de reformulation d'intérêts divers. Ce terme est particulièrement adéquat pour étudier le rôle joué par les pratiques documentaires qui *donnent forme* à des idées vagues et abstraites et les traduisent en programme d'action (voir à ce propos le travail de Florian Charvolin sur « *l'invention de l'environnement* »).

Pierre Lascoumes, *Action publique et environnement*, Que sais-je ?, Paris, Presses Universitaires de France, 2012, 128 p ; F. Charvolin, *L'invention de l'environnement en France*, *op. cit.*

² Charles W. Anderson, « Political Judgment and Theory in Policy Analysis » dans Michael Levy et Edward Portis (eds.), *Handbook of Political Theory and Policy Science*, New York, Greenwood, 1988, p. 183-198.

En suivant les différentes formalisations de la gouvernance de la donnée, trois définitions successives de ce terme sont identifiées :

- la gouvernance comme processus décisionnel des données à ouvrir,
- puis comme administration des données métropolitaines
- et enfin comme politique publique de la donnée métropolitaine.

Ces différentes traductions marquent le passage d'une gouvernance de la donnée perçue comme instances organisationnelles à une gouvernance de la donnée envisagée comme programme d'action publique.

A. La gouvernance comme processus décisionnel de l'ouverture des données

A l'automne 2012, lorsque j'intègre la communauté urbaine de Lyon, la première mission que me confie mon responsable est la « gouvernance de l'ouverture des données ». Profane dans le milieu des collectivités territoriales, je n'ai alors aucune idée de ce que recouvre le terme de « gouvernance de l'ouverture des données » et encore moins de la forme que celle-ci doit prendre. En fouillant dans les archives informatiques, je découvre que la question de la gouvernance est soulevée dès le début du projet *open data*.

Afin d'éviter tout risque « d'interruption de la dynamique initiée avec le projet c'est-à-dire mettre en place une plateforme sans organisation pérenne pour faire vivre les différents enjeux », il est préconisé de « travailler tout au long du projet sur le dispositif de gouvernance future »¹.

Le rôle de la gouvernance semble ainsi de pérenniser la politique d'ouverture des données en l'inscrivant dans l'organisation de la communauté urbaine. Clément, mon responsable hiérarchique, me confirme cette orientation, en me précisant les trois volets qu'elle doit couvrir : « *l'animation interne (promouvoir la démarche et la publication de nouvelles données), l'animation externe (faire connaître le service auprès des communes et des partenaires et faire contribuer au service) et la mise en place d'un comité stratégique de*

¹ Source : Document interne « Présentation du projet *open data* en Commission de validation des études » (9 février 2012)

données »¹. Un échange ultérieur avec François, le second chef de projet *open data* clarifie ce dernier point énigmatique.

Le Comité Stratégique de la donnée assure la cohérence de l'ensemble des données publiées. Par exemple, un partenaire souhaite obtenir de la direction de la voirie l'ensemble des pistes cyclables et de la direction de l'eau la liste des bornes d'eau potable pour créer un service innovant associant les deux jeux de données. La direction de la voirie met à disposition son jeu de données en open data, mais, la direction de l'eau, pour des raisons stratégiques, ne souhaite pas mettre à disposition son jeu de données. Dès lors, c'est au CoStrat de décider de la pertinence et de l'intérêt d'ouvrir cette donnée en fonction du service que souhaite mettre en place le partenaire, pour « obliger » la direction de l'eau à publier également son jeu de données².

Le « comité stratégique de la donnée » est une instance, qui doit permettre à la collectivité de se « *donner la capacité de juger régulièrement, une ou deux fois par an, des nouveaux lots de données à mettre à disposition* »³. La gouvernance de l'ouverture des données est l'élaboration d'une instance de réflexion et de débat sur les données à diffuser. Le chantier qui m'est confié est donc de désigner les acteurs impliqués dans la prise de décision et de définir les règles et les procédures de leur mode d'interaction.

En janvier 2013, je propose alors un schéma de gouvernance de l'ouverture des données qui vise à formaliser à la fois le schéma de prise de décision vis-à-vis des données à ouvrir, à identifier les différents acteurs associés à ces décisions et à définir leurs rôles.

¹ Source : Journal de terrain, réunion avec Clément (19 novembre 2012)

² Source : Journal de terrain, réunion avec François (22 novembre 2012)

³ Source : Interview de Philippe Vial, directeur des systèmes d'information, Millénaire3, 30 novembre 2012, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (Consulté le 15 décembre 2015)

Gouvernance : Organisation SmartData

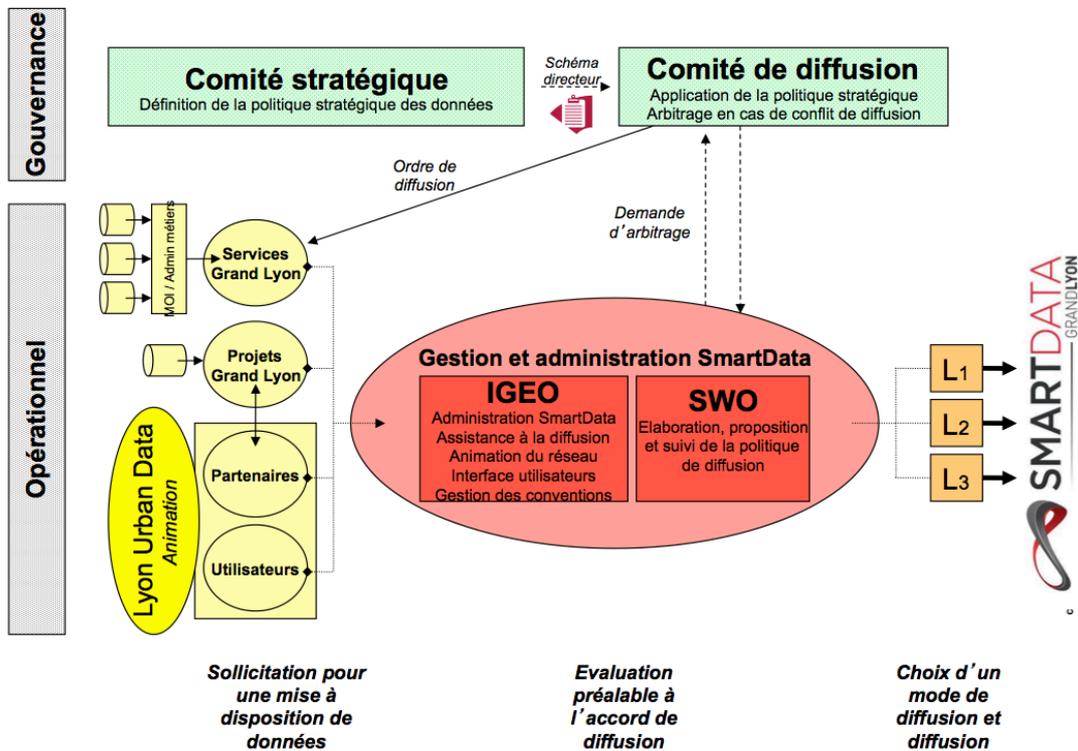


Illustration 30 : Schéma de gouvernance de l'ouverture des données envisagé¹

Ce schéma de gouvernance s'articule autour de trois étapes du processus de mise à disposition des données : la sollicitation d'une donnée à diffuser, l'évaluation de l'intérêt à la diffuser et le choix des modalités de diffusion. L'originalité de ce schéma réside dans la création de deux instances de gouvernance : un comité stratégique, composé d'élus et de représentants des directions, chargé de définir les axes stratégiques de la politique d'*open data* ; et un comité de diffusion, composé du management stratégique de la collectivité, qui arbitre en cas de conflits sur la diffusion d'une donnée en fonction de ces axes stratégiques.

À la vue de ce schéma, la gouvernance de l'ouverture des données correspond à l'organisation du processus de prise de décision des données à diffuser, en identifiant les acteurs qui doivent y prendre part et en déterminant les rôles de chacun. Pour assurer une

¹ Source : Document interne – Présentation au Copil (21 janvier 2013)

cohérence de la politique d'ouverture des données, il est préconisé de créer des instances transversales spécifiquement en charge de cet enjeu.

B. Administrer les données métropolitaines

A l'automne 2013, alors que le schéma décisionnel de « gouvernance de l'ouverture des données » n'a pas encore été mis en œuvre, un nouveau directeur de la direction des systèmes d'information (DSIT) est nommé. Il relance ce chantier, en demandant à quelques agents de sa direction de travailler sur la « gouvernance de la donnée ». Cette deuxième phase voit émerger une nouvelle formalisation de la gouvernance de la donnée : il ne s'agit plus de définir un processus de prise de décision, mais d'administrer les données du territoire.

Loin de s'imposer d'elle-même, cette nouvelle formalisation de la gouvernance de la donnée est le fruit de plusieurs mois d'échanges entre le directeur de la DSIT, un urbaniste du système d'information et les membres de l'équipe du projet *open data* (auquel je participe). Lors de notre première réunion, les incertitudes sur la gouvernance de la donnée sont grandes. Comme je le retranscris alors dans mon journal de terrain, personne ne sait véritablement ce que recouvre cette notion.

François : J'aimerais identifier ce que l'on doit traiter parce que c'est pas clair dans ma tête.

Dimitri : je sais pas exactement l'objet de notre réunion [Il dessine un schéma présentant une chaîne de la donnée comportant trois parties : son intégration dans le SI du Grand Lyon, sa gestion dans le SI et sa diffusion via la plateforme *open data*]: c'est plutôt ça, plutôt ça, ou plutôt ça ?

François : Pour moi, ce n'est pas ça [la diffusion], parce que ça, on l'a déjà cadré [avec les licences de mise à disposition]. C'est pas ça [la gestion des données du SI], point d'interrogation. Moi, je dirai, c'est plutôt ça [l'intégration des données dans le système d'information du Grand Lyon].

Dimitri : Pour moi ce n'est qu'une partie de la gouvernance, c'est la gouvernance du flux, mais pas la gouvernance de la donnée elle-même, la problématique sur la donnée de référence. [...]

François : Alors qu'est-ce qu'on met derrière gouvernance de ces données [de référence]?

Dimitri : Savoir comment elle est utilisée, comment on la mutualise, comment on la diffuse ou pas dans smartdata, est-ce qu'elle est propre quand elle arrive dans smartdata, est-ce qu'on la nettoie dans smart data, c'est toutes ces problématiques là, qui en propriétaire, qui en est responsable,

François : quand on dit gouvernance c'est que autour de la mise à disposition de cette donnée ?

Dimitri : Non, non, non ! Y'a la donnée elle même, y'a la mise à disposition.

Antoine : Il y a trois gouvernances : externe, à la frontière du Grand Lyon pour intégrer les données, interne Grand Lyon pour connaître nos données, à nouveau externe avec la diffusion et les réutilisations.¹

Comme le souligne ce dialogue, aucun des participants ne sait précisément ce que recouvre le terme de « gouvernance de la donnée ». Plutôt, chacun a sa propre perception de ce qu'est la gouvernance de la donnée que l'on tente d'éclaircir dans des échanges ultérieurs. Pour Dimitri, urbaniste des systèmes d'information, la gouvernance de la donnée repose sur la connaissance du cycle de vie de la donnée par la Direction des systèmes d'information. « *L'objectif est de maîtriser le SI et le patrimoine informationnel pour le valoriser. La DSIT gère aujourd'hui le patrimoine applicatif, mais pas le patrimoine informationnel*². » Pour Pauline, géomaticienne et responsable de la gestion de la plateforme *open data*, l'enjeu de la gouvernance des données est d'établir un cadre conventionnel clair d'acquisition et de diffusion de données. Enfin, pour ma part, l'enjeu de la gouvernance des données est de créer des arènes de débat et de décision des choix politiques à chacune des étapes de la chaîne de la donnée (production, gestion, analyse, diffusion, restitution, etc.).

Un an d'échanges entre ces différents acteurs sera nécessaire pour dépasser cette ambiguïté lexicale, réunir les différents univers de sens et parvenir à élaborer un consensus sur le terme de gouvernance de la donnée. En décembre 2014, le directeur des systèmes d'information présente en comité de direction générale un document synthétique de ce travail intitulé « Gouvernance de la donnée »³. Par son travail de « mise en rapport⁴ » des différentes significations de la gouvernance des données, cette présentation marque une formalisation stabilisée de la gouvernance des données comme administration des données métropolitaines. Elle clarifie également le rôle de la Métropole de Lyon, et de sa direction des services informatiques, dans cette gouvernance.

¹ Source : Journal de terrain (14 octobre 2013)

² Source : Journal de terrain, réunion Gouvernance des données (4 février 2014)

³ Source : Document interne « Gouvernance des données » (12 décembre 2014)

L'argumentation développée dans les deux paragraphes suivants s'appuie sur l'analyse de ce document.

⁴ Florian Charvolin met en avant cette double dimension de la mise en rapport « *qui signifie à la fois rapprochement d'énoncés différents et élaboration d'un document papier* ». Le rapport est ainsi à la fois document et tissu de relations quand il réussit à faire tenir ensemble différentes significations. F. Charvolin, *L'invention de l'environnement en France, op. cit.*, p. 34.

Le document débute par une présentation de l'enjeu de la donnée tel qu'il est problématisé par les agents de la direction des systèmes d'information. Les données sont perçues comme créatrices de valeur pour le développement économique, pour la « modernisation de l'action publique » et pour la « transparence et l'implication citoyenne ». A contrario, des risques sont identifiés : l'appropriation des données par des acteurs privés et la formation de monopoles, la perte d'indépendance et de neutralité ainsi que les risques juridiques autour des données personnelles. Pour traiter à la fois de ce potentiel de valeur et de ces risques, la direction des systèmes d'information propose de créer un « service public de la donnée¹ », dont la finalité principale est de mettre en place une gouvernance des données du territoire métropolitain. La Métropole de Lyon est identifiée comme l'acteur légitime pour organiser ce service public de la donnée en organisant un réseau d'acteurs sur l'agglomération.

La « gouvernance de la donnée » prend dorénavant la forme d'une administration des données de la Métropole de Lyon, mais également des différents acteurs du territoire (producteurs, utilisateurs, etc.). Elle s'articule autour de plusieurs axes : une réflexion sur les données à produire et à acquérir, une gestion des données (connaissance du patrimoine informationnel, maîtrise de la qualité et normalisation des données), la mise en place de « leviers de fluidité » pour faciliter la circulation des données et enfin la création de valeur à partir des données en développant les usages. Ces principes généraux forment les grands axes de cette deuxième formalisation de la « gouvernance de la donnée » dont l'objet n'est plus d'établir un schéma de prise de décision sur l'ouverture des données, mais d'administrer la donnée du territoire métropolitain en intervenant sur l'ensemble de la chaîne de la donnée.

Pour mettre en œuvre ce programme, une recomposition des relations entre l'ensemble des acteurs de la chaîne de la donnée est présentée au profit de la Métropole de Lyon qui devient « *l'acteur central* » définissant « *les standards et les bonnes pratiques* ». La gouvernance de la donnée prend la forme d'un « *réseau d'acteurs* » dont la Métropole de Lyon est l'« *autorité organisatrice* ». Afin d'animer ce réseau, des instances de pilotage,

¹ Cette notion de « service public de la donnée » est floue et ambiguë : ce qu'elle recouvre n'est jamais véritablement explicité.

« techniques » et « politiques », regroupant des acteurs internes et externes à la Métropole, sont proposées.

Cette prétention du Grand Lyon à devenir « *autorité organisatrice* » des données à l'échelle de la métropole se traduit par une nouvelle définition des données. L'objet de la gouvernance n'est en effet plus uniquement les données de la collectivité, mais les données du territoire métropolitain, qu'elles soient publiques ou privées. Pour justifier cet élargissement de l'objet de la gouvernance, le directeur des systèmes d'information met en avant la notion de « données d'intérêt général¹ ».

Aujourd'hui, on n'est pas sur un schéma de distinction avec d'un côté des données publiques, et de l'autre côté des données privées, avec une approche très réglementaire et juridique. On est plutôt sur une approche politique de la donnée en disant : il y a des données qui ne servent pas l'intérêt général, et il y a des données qui servent l'intérêt général. Ça veut dire que l'on a un enjeu à nous positionner, non pas sur la gouvernance de la donnée publique, mais sur la gouvernance de la donnée d'intérêt général. [...] Et donc potentiellement à contraindre des opérateurs privés à ouvrir des données du fait qu'elles sont considérées d'intérêt général. Ça questionne sur les données type Uber par exemple. Pour pouvoir exercer une politique dite de mobilité, on a besoin que l'ensemble des acteurs qui génèrent de la donnée d'intérêt général l'ouvre de manière à permettre la création de services multimodaux que l'on ne pourrait pas faire si on considère que c'est de la donnée privée. Cette vision là, données d'intérêt général versus données publiques/données privées, paraît être une avancée assez intéressante. [...] On pousse en avant l'idée d'un service public de la donnée et la prise en compte de la donnée comme une politique publique à part entière. Auquel cas, on aurait un service public de la donnée, et qui n'est pas que de la donnée publique, mais qui porte bien la notion de données d'intérêt général².

¹ Cette notion de « données d'intérêt général » est issue du Rapport Jutand sur l'ouverture des données de transport, qui en propose la définition suivante : « Une information d'intérêt général (IIG) est une information produite dans le cadre de services au public dont l'ouverture est considérée d'intérêt général, car elle rend possible la mise en place de nouveaux services à l'intention du public ». (Source : *Ouverture des données de transport*, Paris, Rapport remis au secrétaire d'État chargé des Transports, de la Mer et de la Pêche, 2015, p. 72.)

Le Conseil national du numérique s'est prononcé en défaveur de la création de ce statut estimant que la notion « par définition de portée trop large et floue », serait « trop complexe pour être mobilisable simplement ». (Source : *Ambition numérique. Pour une politique française et européenne de la transition numérique*, Paris, Conseil national du numérique, 2015, p. 153)

Le projet de loi pour une République numérique adopté par le gouvernement le 9 décembre 2015 introduit toutefois cette notion de « données d'intérêt général » afin de favoriser « l'ouverture des données issues de personnes publiques et privées, titulaires de délégations de service public ou dont les activités sont subventionnées par la puissance publique, et en permettant un accès simplifié de la statistique publique à certaines bases de données privées ». (Source : Projet de loi pour une République numérique, <http://www.republique-numerique.fr/media/default/0001/02/ce21a30ba6d31b99c71311438a172e3c547c9dca.pdf> (Consulté le 15 décembre 2015).

² Source : Entretien avec Tom, directeur de la DINSI, Grand Lyon (18 mars 2015)

Cette traduction de la donnée d'une conception « juridique » vers une conception « politique » légitime la prétention du Grand Lyon à gouverner les données de l'ensemble des acteurs qui agissent sur le territoire métropolitain. La distinction juridique données publiques / données privées limitait l'action du Grand Lyon aux données de la communauté urbaine. En créant la notion de « donnée d'intérêt général », dont la définition reste volontairement floue¹, il élargit la légitimité de son action à l'ensemble des données relevant des politiques métropolitaines qu'elles soient détenues par des acteurs publics ou privés. Cette nouvelle conception de la donnée permet à la Métropole de Lyon de s'insérer au cœur de la chaîne des données métropolitaines en se positionnant en situation d'intermédiaire entre producteurs et utilisateurs.

*

* *

Ce document de présentation de la gouvernance de la donnée parvient à agréger les différentes orientations qui se sont préalablement manifestées. Elle formalise la gouvernance comme une réponse au besoin de connaître et de maîtriser le système d'information, comme un outil pour réguler l'acquisition et la diffusion de données, et comme un ensemble d'instances pour en discuter les enjeux et la finalité. En stabilisant la définition de la gouvernance de la donnée, ce document doit permettre à la Métropole de Lyon de construire une politique publique de la donnée.

C. Une politique publique de la donnée métropolitaine ?

La troisième formalisation de la « gouvernance de la donnée » intervient quelques mois plus tard, au printemps 2015, lorsque ces principes généraux sont traduits en programme d'actions à mener². Il s'agit alors « *d'affirmer le positionnement de la Métropole comme*

¹ Cette ambiguïté facilite le soutien d'acteurs aux intérêts divers qui peuvent chacun y trouver une interprétation acceptable.

² Cette sous-partie s'appuie sur l'analyse de divers documents internes produits au printemps et à l'automne 2015 et d'un entretien avec l'administratrice générale des données du Grand Lyon.

porteur du service public de la donnée sur son territoire¹ » en suivant trois axes prioritaires.

Le premier axe vise à inventorier les données métropolitaines en les identifiant et les qualifiant par la création d'un catalogue de la donnée métropolitaine. Cet inventaire doit assurer la « *bonne gestion du patrimoine informationnel existant en garantissant l'accès, la normalisation, l'interconnexion et la disponibilité²* » des données du territoire.

Le deuxième axe souhaite positionner la Métropole comme « Autorité organisatrice de la donnée » en centralisant et en assurant la diffusion des données de certains secteurs d'action publique. Trois secteurs prioritaires ont été identifiés : la mobilité, l'énergie et le *care*. Il s'agit de poursuivre le processus de centralisation et de mise en relation de l'ensemble des données du secteur de la mobilité entamé dans le cadre du projet Optimod et de l'étendre à deux autres secteurs d'action publique.

Le troisième axe vise à accompagner l'appropriation et la réutilisation des données en développant des actions de sensibilisation et de médiations à la donnée auprès de différents acteurs, en animant un écosystème d'utilisateurs et en régulant l'usage des données dans le respect de l'intérêt général.

Ce programme de gouvernance de la donnée vise quatre cibles d'acteurs : les « producteurs de données internes à la collectivité », les « *utilisateurs internes à la collectivité* », les « *producteurs externes à la Métropole (partenaires, communes, acteurs des filières économiques prioritaires et prestataires)* » et les « *utilisateurs externes (monde académique, Tubà, communes, partenaires, acteurs des filières économiques prioritaires)*³ ».

La transposition des principes généraux d'administration de la donnée en programme d'actions marque une nouvelle formalisation de la « gouvernance de la donnée ». Les énoncés

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² Source : Document interne « Les données métropolitaines, une nouvelle dimension du territoire » (24 août 2015)

³ Source : Document interne « Les données métropolitaines, une nouvelle dimension du territoire » (24 août 2015)

restent généraux et les objectifs à atteindre sont peu précisés. Mais la gouvernance de la donnée est désormais une politique publique en voie de constitution avec des enjeux à résoudre, un programme d'action et des cibles sur lesquelles agir.

*

* *

En retraçant la formalisation de la « gouvernance de la donnée » au sein du Grand Lyon, on observe à quel point celle-ci est l'objet de traductions successives. Initialement perçue comme l'élaboration d'un processus et d'instances de décision des données à mettre à disposition, la gouvernance de la donnée est devenue une ébauche de politique publique dotée de principes généraux à suivre et d'un programme d'actions à mettre en œuvre. Ce n'est plus l'*open data* qui est l'objet de la gouvernance, mais la donnée, dont la diffusion n'est qu'une des modalités. Comme me le confie un de mes interlocuteurs : « *la problématique, ce n'est plus l'open data, c'est la data*¹ »

Ces différentes traductions de la « gouvernance de la donnée » font évoluer les acteurs associés et le périmètre de l'enjeu à gouverner. Alors qu'initialement les données sont restreintes aux données « à ouvrir », c'est l'ensemble des données de l'institution communautaire et de l'espace métropolitain qui entrent finalement dans le spectre de cette gouvernance. Dès lors, si auparavant seuls les acteurs internes à l'organisation étaient visés (producteurs de données), la cible de la gouvernance s'étend désormais à l'ensemble de la chaîne de la donnée : producteurs, diffuseurs, utilisateurs, etc., que ces acteurs appartiennent ou non au Grand Lyon. Enfin, les personnes désignées pour prendre en charge cette gouvernance de la donnée évoluent également. Initialement, le schéma de gouvernance regroupe une pluralité d'acteurs (élus, représentants des directions du Grand Lyon, service Web et Organisation, service de l'information géographique). La traduction de la gouvernance attribue désormais à la direction des systèmes d'information un poids majeur dans sa mise en œuvre. Cela se traduit dans l'institutionnalisation de la « gouvernance de la donnée ».

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

II. Institutionnaliser la gouvernance de la donnée

Après avoir étudié comment la gouvernance des données a été construite comme catégorie de politique publique, la seconde partie de ce chapitre s'intéresse à l'institutionnalisation de la gouvernance de la donnée. L'institutionnalisation est entendue comme l'intégration de cet enjeu de gouvernance de la donnée au sein de l'organisation administrative de la Métropole de Lyon¹. Cette institutionnalisation de la gouvernance des données s'effectue par le biais de deux processus : l'autonomisation de cet enjeu « donnée » vis-à-vis des secteurs traditionnels de politique publique et la professionnalisation d'acteurs administratifs spécifiquement dédiés à ce sujet. Pour étudier ces processus, quatre points sont abordés :

- l'impossible institutionnalisation de l'ouverture des données,
- la recomposition de la direction des systèmes d'information,
- la reconfiguration du service d'informations géographiques
- et la création du poste d'administrateur général de la donnée.

A. Le refus d'autonomiser la politique d'ouverture des données

La première formalisation de la gouvernance de la donnée a donné lieu à l'élaboration d'un schéma organisationnel. Alors que la décision de mise à disposition des données est effectuée au sein de chacune des directions métiers, l'équipe projet souhaite créer des instances politiques transversales spécifiquement dédiées à ce sujet. Dans une note au directeur général, le chef de projet *open data* justifie cette volonté d'autonomiser la politique d'ouverture des données des politiques sectorielles.

L'ouverture des données soulève des enjeux politiques sur lesquels les élus peuvent être sollicités par les citoyens. Elle requiert à ce titre une gouvernance spécifique afin de définir les grandes orientations que doit suivre la politique de diffusion des données du Grand Lyon et d'assurer une cohérence entre les données des différents domaines de compétence de l'institution. [...] L'ouverture des données nécessite une transversalité entre les différentes politiques publiques du

¹ Il s'agit ainsi d'une définition très restrictive de l'institutionnalisation limitée au processus de création d'une organisation formelle en charge de la mise en œuvre de cette politique publique.

Grand Lyon qui n'existe pas dans les instances actuelles. La pluridisciplinarité et le dépassement de la vision en silos des commissions actuelles est pour nous un des facteurs de réussite de notre projet¹.

La direction générale refusera la création d'une commission d'instances spécifiques. L'ouverture des données n'est pas perçue comme un enjeu autonome nécessitant la mise en place d'instances transversales. Il est préconisé que ce sujet soit débattu, de manière ponctuelle, au sein des commissions politiques existantes.

Malgré une mobilisation de l'équipe du projet *open data*, le schéma décisionnel initialement imaginé ne sera jamais mis en place. Ce refus révèle les visions divergentes qui s'opposent vis-à-vis de la donnée entre d'un côté la volonté d'en faire un enjeu propre et de l'autre le souhait d'intégrer ces questions dans les politiques sectorielles existantes. La première formalisation de la gouvernance de la donnée, centrée autour des problématiques de sa diffusion, ne parviendra ainsi pas à s'institutionnaliser dans l'organisation de la communauté urbaine.

B. Des systèmes d'information au numérique : une autonomisation des services informatiques

La création de la Métropole de Lyon le 1^{er} janvier 2015 s'est accompagnée d'une refonte de l'organisation de l'administration. Parmi les changements principaux figure le transfert de l'ancienne direction des systèmes d'information et des télécommunications (DSIT) de la délégation générale aux ressources vers une délégation au développement économique à l'emploi et aux savoirs (DDEES). Ce passage marque une transformation du périmètre de la direction des services informatiques – renommée direction à l'innovation numérique et aux systèmes d'information (DINSI) – qui regroupe l'ancienne DSIT et la mission numérique auparavant rattachée à la délégation générale au développement économique et internationale (DGDEI), ainsi que le centre d'innovation numérique Erasme², issu du département du Rhône³. La création de la DINSI marque une évolution de son rôle de

¹ Source : Document interne « Note sur la gouvernance de l'ouverture des données » (28 août 2013)

² Erasme est un living lab créé en 1998 avec pour mission de mettre en place des méthodes d'innovation ouverte pour susciter de nouveaux usages du numérique pour l'éducation, la culture et les services à la personne.

³ En termes d'effectif, la DSIT comptait environ 120 agents, la DINSI en compte 160.

la gestion des systèmes d'information au développement d'une politique du numérique. Celle-ci doit « *faire du numérique un levier de transformation de la métropole (moderniser l'administration, expérimenter et innover pour créer de la richesse et de l'emploi, administrer électroniquement et associer l'utilisateur dans une administration ouverte)*¹ ». La création de la DINSI traduit le souhait de considérer les services numériques et le système d'information comme une politique publique à part entière. « *Depuis, le printemps 2013, ça fait quasiment deux ans que je pousse le fait qu'il faut que l'on aille vers une vision qui soit systémique, de ne pas aller vers d'un côté le numérique pour le dehors, et de l'autre côté l'informatique pour le dedans*². » La DINSI se positionne, à la fois comme une ressource pour les services internes et ouverte vers les acteurs externes à la Métropole. A la différence de l'introduction de l'informatique dans les administrations publiques étudiée par Pierre-Yves Baudot³, la bannière du « numérique » permet aux « informaticiens » de se présenter non seulement comme les pilotes de la réorganisation des méthodes de travail de l'administration, mais également comme les porteurs de nouveaux services à destination des usagers. En ne gérant plus uniquement les systèmes d'information de l'institution, la DINSI s'autonomise des autres directions de la Métropole pour mener sa propre politique publique du numérique⁴.

Si elle n'est pas la cause directe de cette évolution de la DSIT à la DINSI, la gouvernance de la donnée participe indubitablement au repositionnement stratégique des services informatiques. « *La réorganisation a permis à la fois de repositionner la donnée au cœur de l'organisation et au cœur de la stratégie* »⁵ indique le directeur de la DINSI. La maîtrise de la donnée est en effet une ressource importante pour participer au pilotage de l'administration et mettre en œuvre des politiques de développement comme le souligne cet interlocuteur :

¹ Source : Document interne, « Projet de délégation Dév Eco, Emploi et Savoirs » (16 avril 2014)

² Source : Entretien avec Tom, directeur de la DINSI, Grand Lyon (18 mars 2015)

³ Pierre-Yves Baudot, « L'informatique aux informaticiens. Sélection des acteurs, définition de l'instrument et politisation de l'informatique administrative (France, 1966-1975) » dans Gwenola Le Naour et Gilles Massardier (eds.), *L'action publique sélective*, Paris, LGDJ, 2013, p. 182-196.

⁴ Cette autonomisation se manifeste par l'intégration en 2015 d'une politique « Numérique et ville intelligente » dans le référentiel établissant les 23 politiques publiques de la Métropole de Lyon.

⁵ Source : Entretien avec Tom, directeur de la DINSI, Grand Lyon (18 mars 2015)

Le directeur de la DSIT a vite vu qu'il y avait un intérêt stratégique à positionner sa direction comme gestionnaire de la donnée, puisqu'on sait que la gestion de l'information dans une collectivité, c'est quelque chose qui peut vite devenir central. Notamment, si on a des velléités pour aller vers une direction de l'innovation et des systèmes d'information, comme c'est le cas aujourd'hui, donc d'aller plutôt vers des problématiques et des sujets de la ville intelligente et d'une direction plus orientée vers du développement. [...] Gérer la donnée de manière centralisée, c'est une façon de se positionner comme un passage obligatoire. Stratégiquement, c'est quand même pas bête d'être au cœur de la donnée¹.

La gouvernance de la donnée est un moyen pour la direction des services informatiques de s'autonomiser des directions « métiers » de la collectivité et de mener sa propre politique du numérique. Ce nouveau positionnement de la direction des services informatiques d'une fonction support à une fonction de pilotage d'une politique publique se manifeste dans la reconfiguration de son service d'information géographique qui est dorénavant chargé de l'administration des données métropolitaines.

C. Du service de l'information géographique à l'administration des données métropolitaines

Durant les trois années de mon étude au sein de la communauté urbaine de Lyon, le service de l'information géographique a subi un renouvellement important de ses missions. De producteur des données de référence, il est devenu gestionnaire de la plateforme de diffusion de données, puis administrateur des données métropolitaines.

L'*open data* a contribué à renouveler l'organisation du service d'information géographique du Grand Lyon. En effet, ce service qui était en charge de la production de données géographiques de référence et de leur diffusion auprès des directions et des partenaires de la communauté urbaine, s'est vu confier la gestion quotidienne de la plateforme de diffusion des données. Cette nouvelle mission a transformé le quotidien des agents de ce service. Ils étaient auparavant chargés de la diffusion des données géographiques de référence par l'intermédiaire d'un guichet de vente. Ils doivent dorénavant assurer la bonne diffusion des données de l'agglomération par l'intermédiaire d'une infrastructure informatique. Ils ne vendent plus de la donnée à des « clients », mais ils la diffusent à des « utilisateurs » comme le souligne l'ancien responsable de ce service :

¹ Source : entretien avec Clément, chef de projet *open data*, Grand Lyon (17 février 2015)

[L'open data] ça nous a obligé à faire évoluer nos pratiques administratives et juridiques. Dans le fonctionnement courant, l'unité diffusion et géoservices est en très très forte évolution, parce qu'en espace de quelques années, on a des gens qui étaient profilés pour accueillir du public derrière un guichet et vendre de la donnée. On ne vend plus de données, on n'accueille plus de public, et on n'a plus de guichet. On a par contre à gérer de la donnée sur le portail SmartData. Il y a eu des changements de profils, il y a une évolution des métiers et des personnes qui sont en place¹.

Alors qu'auparavant, le service Igéo n'assurait que la diffusion des informations géographiques de référence qu'il produisait, il est dorénavant en situation d'intermédiaire entre les producteurs et les réutilisateurs de données. Le périmètre de son domaine d'intervention s'est élargi pour embrasser les données, géographiques ou non, de l'ensemble des directions de la communauté urbaine. Cette évolution transforme les relations du service Igéo avec ces acteurs qui étaient auparavant autonomes dans la gestion de la diffusion de leurs données. Pour la direction des systèmes d'information, cette situation l'empêchait de maîtriser les données du système d'information. *« Aujourd'hui la DSIT n'a aucune visibilité sur ces échanges : chaque direction gère avec les interlocuteurs ses échanges de données. Or, c'est un aspect problématique en termes de gouvernance du système d'information, sur lequel [le directeur de la DINSI] aimerait retrouver une prise au sein de la DSIT². »* En centralisant la gestion de la diffusion au sein de son service de l'information géographique, la DSIT acquiert une meilleure connaissance de la diffusion des données métiers des directions. *« On ne le voyait pas avant. On n'a jamais géré la diffusion de données métiers, et on n'a jamais eu la capacité de savoir vraiment ce qui se diffusait. La voirie, l'eau, l'urbanisme, ça ne passait pas du tout par nous. Chacun était responsable et c'était complètement invisible pour nous »³.* La maîtrise de la diffusion des données de l'institution était ainsi restreinte. En gérant l'infrastructure technique et humaine de diffusion des données du Grand Lyon, la DSIT parvient à avoir une « prise » sur les données de l'institution. En situation d'intermédiaire entre producteurs et réutilisateurs, le service de l'information géographique tente de se positionner comme un « point de passage obligé » de la politique de la donnée métropolitaine. Grâce à la gestion de la plateforme *open data*, le service Igéo se retrouve au cœur de la gouvernance de la donnée métropolitaine.

¹ Source : entretien avec Quentin, responsable du service Igéo, Grand Lyon (15 janvier 2014)

² Source : Journal de terrain, Réunion Gouvernance de la donnée (17 septembre 2013)

³ Source : Entretien avec Etienne, administrateur de données, Grand Lyon (26 novembre 2013)

La création de la DINSI formalise ce nouveau rôle d'administrateur de la donnée. Le service de l'information géographique (Igé) est renommé « Géomatique et données métropolitaines ». Surtout, il voit ses missions évoluer puisque son rôle n'est plus uniquement de gérer la diffusion des données, mais également de piloter la politique publique de la donnée en définissant des orientations stratégiques qui s'imposent aux producteurs.

Aujourd'hui, on est administrateur des données fédérales, c'est-à-dire les données dont toutes les directions ont besoin. C'est ce qu'on a avec Igéo qui gère les données de référence pour l'ensemble des métiers. Pour toutes les autres données, les directions gèrent en propre, de manière verticale. L'enjeu est de devenir demain administrateur fédéral de données. C'est-à-dire de passer à un niveau au dessus en définissant et mettant en œuvre des orientations stratégiques sur la donnée pour l'ensemble des directions. La diffusion des données sur SmartData nous a conduit à entrer dans cette gestion des données métiers en harmonisant les processus. Aujourd'hui, il faut aller plus loin et penser en amont la donnée des directions¹.

Tout en poursuivant ses missions initiales de production de données de référence, le service de l'information géographique intègre dorénavant cet enjeu de pilotage de la politique publique de la donnée². De la diffusion, il remonte en amont la chaîne de la donnée pour définir aux producteurs des orientations à suivre. Dans cette optique, un nouveau poste est créé en son sein : celui d'administrateur général des données³.

D. Une nouvelle fonction : l'administratrice générale des données

La volonté de la DINSI d'être au cœur de la gouvernance de la donnée métropolitaine conduit à la création d'une nouvelle fonction d'administrateur général des données. Cette agent a pour rôle de piloter la politique publique de la donnée métropolitaine. Si la création de ce poste marque une indéniable professionnalisation de la politique de la donnée, le contour des missions de cet administrateur général des données reste flou.

¹ Source : Journal de terrain, Réunion avec Tom, directeur de la DINSI, Grand Lyon (24 mars 2014)

² Le pilotage de la gouvernance de la donnée par le service de l'information géographique s'explique notamment par l'expertise développée par les géomaticiens dans l'administration des données géographiques qu'ils étendent dorénavant à d'autres types de données.

³ Le terme d'administrateur marque la filiation avec le milieu de la géomatique. En effet, le poste d'administrateur de donnée existait auparavant au Grand Lyon, mais il était réservé aux responsables des données géographiques. L'adjectif général qu'on lui ajoute traduit le nouveau positionnement du service Igéo dans un rôle stratégique de pilotage de la « gouvernance » de l'ensemble des données métropolitaines.

Le poste d'administrateur général des données s'inspire directement des *Chief Data Officers* (CDO) nommés dans certaines entreprises et administrations municipales nord-américaines¹. La circulation de cette fonction des mondes privés et nord-américains vers les administrations publiques françaises a débuté par la nomination en mai 2014 d'Henri Verdier, alors responsable de la mission gouvernementale Etalab, comme administrateur général des données de l'État. Quelques mois plus tard, l'administration du Grand Lyon acte le recrutement d'un « administrateur général des données » qui est placé au sein du service « Géomatique et données métropolitaines » de la récente DINSI. La définition de ce nouveau poste au sein de l'administration découle directement des problématiques et des enjeux soulevés lors des travaux préparatoires à la mise en place d'une gouvernance de la donnée.

Ce que l'on avait imaginé faire [pour élaborer la fiche de poste du CDO], c'est repartir de la démarche que vous avez eu sur la gouvernance de la donnée. En gros, ce qui est ressorti comme les missions sur la gouvernance de la donnée, les transcrire dans un profil de poste².

Les missions définies dans la fiche de poste de l'administrateur général de données révèlent la volonté de la DINSI d'occuper une place centrale dans la chaîne de la donnée. A partir d'un inventaire des données, l'administrateur devra faciliter la circulation et l'usage des données tant au sein de l'administration que vers des acteurs externes.

1. demande aux services et aux acteurs du territoire l'**inventaire des données** qu'ils produisent, reçoivent ou collectent,
2. **facilite leur circulation** au sein de l'administration comme à destination des chercheurs, entreprises et citoyens dans le respect de la protection de la vie privée et des secrets protégés par la loi,
3. s'implique dans la **production** de données clés,
4. **contribue à la qualité** des données produites,
5. **coordonne** l'action en matière d'inventaire, de gouvernance, de production, de circulation et d'exploitation des données,
6. **stimule la diffusion** des nouvelles méthodes d'action fondées sur la donnée : approches big data, allocation des ressources publiques, stratégies fondées sur la donnée, etc.
7. **conduit des expérimentations** sur l'utilisation des données pour renforcer l'efficacité des politiques publiques, contribuer à la bonne gestion des deniers publics et améliorer la qualité des services rendus aux usagers,

¹ L'administratrice générale des données se nomme elle-même comme le « *chief data officer* » de la Métropole de Lyon.

² Source : Entretien avec Etienne, administrateur de données, Grand Lyon (18 février 2015)

8. propose une **position métropolitaine** dans les négociations portant sur la politique de la donnée,

9. remet **chaque année un rapport public** sur l'inventaire, la gouvernance, la production, la circulation, l'exploitation des données¹

La description de ses missions souligne le rôle majeur que doit jouer cet administrateur général dans la gouvernance des données de l'institution. Il doit connaître les données produites sur le territoire, s'assurer de leur qualité, encourager leur circulation, définir les données cruciales à produire et maintenir, expérimenter de nouvelles méthodes basées sur l'usage des données, élaborer et défendre une politique métropolitaine de la donnée pour laquelle il devra rendre des comptes annuellement.

Encadré 15 : La fonction de *Chief Data Officer* dans les administrations municipales américaines

La fonction de *Chief Data Officer* a fait son apparition au milieu des années 2000 aux Etats-Unis au sein d'entreprises de l'économie numérique² et du secteur de la banque et des assurances³. Ce poste de responsable des données des organisations a ensuite été adopté au sein d'administrations publiques américaines⁴. En 2012, les municipalités de New York, Philadelphie et Chicago ont été les premières à se doter d'un *Chief Data Officer*, suivies par une dizaine de villes (Los Angeles, Houston, San Francisco, etc.)⁵. Les CDO recrutés sont majoritairement des personnes ayant une formation en informatique et/ou en statistiques⁶. Plusieurs d'entre eux ont également fait partie de l'organisation de *civic hackers* Code for

¹ Source : Document interne – « Gouvernance de la donnée » (Décembre 2014)

² Usama Fayyad, considéré comme l'un des premiers *Chief Data Officer* a été nommé à ce poste en janvier 2004 chez Yahoo.

³ Selon Gartner, le poste de *Chief Data Officer* s'est développé fortement à partir de 2010. En 2013, plus d'une centaine d'entreprises disposaient d'un CDO, le double que l'année précédente. Les CDO sont principalement recrutés dans les secteurs de la banque, de l'assurance et du gouvernement et ils sont situés pour près de 65% d'entre eux aux Etats-Unis.

Source : http://blogs.gartner.com/mark_raskino/2013/11/06/5-facts-about-chief-data-officers (Consulté le 15 décembre 2015)

⁴ En 2010, l'Etat du Colorado est la première administration américaine à se doter d'un CDO.

⁵ Des postes de *Chief Data Officers* ont également été créés ces dernières années au sein des administrations régionales et fédérales.

⁶ Le profil et les missions types des CDO ont été analysés, en décembre 2015, à partir des profils LinkedIn d'une quinzaine de *Chief Data Officers* de municipalités américaines.

America. Les CDO sont le plus souvent rattachés administrativement soit aux services informatiques soit aux départements de l'évaluation et des performances de leur institution¹. Leurs missions principales, plus ou moins accentuées selon les municipalités, consistent à favoriser la circulation des données au sein de l'administration (*data repository*), encourager l'ouverture des données (*open data*), faciliter l'interopérabilité des données de l'institution et développer des programmes d'analyse de données (*analytics*).

En avril 2015, la Métropole de Lyon recrute son administrateur général des données. Comme l'indique un communiqué publié à cette occasion, « Créé au sein du service « Géomatique et Données Métropolitaines », ce poste accompagne la transformation de l'ancien service « de l'Information Géographique » dans son nouveau positionnement. Le rattachement du CDO à cette entité démontre la volonté stratégique de mettre l'expertise acquise sur la gestion et l'administration des données géographiques au service de toutes les données, pour en centraliser la gouvernance². » Dans ce communiqué de presse, les missions de cet administrateur sont précisées. Il est en charge de la définition et de la mise en œuvre de la stratégie de gouvernance des données, missions qu'il doit mener dans une démarche partenariale avec les différents acteurs de l'agglomération (acteurs publics, entreprises, universités, société civile, etc.).

Il s'agit d'élaborer et de faire appliquer une stratégie de gouvernance des données, produites par l'ensemble des entités de la Métropole tout en mobilisant le milieu économique local dans une démarche d'ouverture et de diffusion des données. L'animation de partenariats tout comme le fait de faire entendre la position de la métropole lyonnaise dans les instances nationales et internationales qui traitent de la data, seront des axes forts de l'activité du CDO. L'objectif de l'administrateur des données est enfin de créer un cadre favorable à la contribution des citoyens, usagers-experts, développeurs, chercheurs, associations, startups... pour valoriser les données au travers de développements de services numériques, pour améliorer ainsi en continu le cadre de vie lyonnais, pour offrir au territoire métropolitain de nouvelles opportunités de mises en lumière de ses nombreuses ressources en matière d'innovation³.

¹ A l'exception notable de la ville de New York et de l'Etat français, où les CDO sont rattachés directement au pouvoir exécutif (*Mayor's Office*, Secrétariat général pour la modernisation de l'action publique, administration du Premier ministre).

² Source : « La Métropole de Lyon accueille son « Chief Data Officer » », Communiqué de Presse, 11 mai 2015, <http://decryptageo.fr/la-metropole-de-lyon-accueille-son-chief-data-officer> (Consulté le 17 novembre 2015)

³ Source : « La Métropole de Lyon accueille son « Chief Data Officer » », Communiqué de Presse, 11 mai 2015, <http://decryptageo.fr/la-metropole-de-lyon-accueille-son-chief-data-officer> (Consulté le 17 novembre 2015)

En avril 2015, Léa est nommée à ce poste. Son profil s'inscrit à la croisée des technologies numériques et de la communication puisqu'elle a occupé auparavant successivement les postes de responsable de l'intranet, de l'extranet et du web économique du Grand Lyon. Six mois plus tard, je l'interroge sur ce poste, dont elle ne parvient encore à définir précisément le périmètre, mais qu'elle inscrit dans la continuité de la politique d'*open data*.

Léa : Les axes pour moi, ils sont multiples. Au bout de six mois, j'ai encore du mal à prioriser tout ça. [...] Je vois l'ampleur de la tâche, et il y a encore des choses à caler, je pense que j'y verrai plus clair dans quelques mois.

Enquêteur : quand tu as été recrutée sur ce poste, les missions étaient définies plus ou moins clairement ?

Léa : Elles étaient définies en rapport avec le travail que vous aviez fait sur la gouvernance. Mais, je n'ai toujours pas une fiche de poste. Ça ne m'affecte pas, mais il y a quand même un moment où il faudra qu'on écrive clairement ce qu'un CDO doit faire. Il y a un volet de communication, mais il y a aussi un volet "on structure les choses", on met en œuvre tous les leviers qui permettent l'ouverture de la donnée¹.

Ce poste, récent, suscite de nombreuses incertitudes sur son périmètre. Les échanges que Léa a pu avoir avec des personnes occupant un poste similaire au sein des collectivités de Manchester, de Gand et de Gijón confirme ces questionnements. « *Tous les trois souhaitent travailler sur le contenu de leur poste. Ils se posent plein de questions sur le périmètre de leur poste².* ». Ce poste de *Chief Data Officer* est toutefois intimement lié aux politiques d'ouverture des données³. Autant dans les villes nord-américaines que j'ai étudiées qu'au sein du Grand Lyon, il vise à pérenniser les démarches d'*open data*. Léa est chargé de poursuivre l'ouverture des données, de formaliser les processus de diffusion, de mettre en place des instances de décisions, d'identifier des secteurs d'action publique dont les données sont à ouvrir prioritairement, d'effectuer un travail de lobbying pour défendre les licences de mise à disposition, de dialoguer avec les réutilisateurs. Le CDO doit ainsi institutionnaliser l'*open data* en élaborant une stratégie, en mettant en place les outils, en établissant des processus, etc. Léa présente son rôle comme l'intermédiaire entre les producteurs et les utilisateurs de données. « *Je suis au milieu du circuit, entre les producteurs et les utilisateurs. Je fais le*

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

³ Signe de ce lien étroit, l'administratrice générale des données du Grand Lyon partage d'ailleurs son bureau avec la responsable de la plateforme de diffusion de données.

lien¹ ». Le recrutement du CDO s'inscrit ainsi dans le prolongement des politiques d'*open data*, tout en les questionnant pour considérer stratégiquement la *data*.

On est dans une seconde phase de l'*open data*. On a tous mis en œuvre des stratégies pour ouvrir les données, des plateformes. Et là en fait, on est dans une phase où il faut se poser les bonnes questions. Il faut se poser des questions vraiment stratégiques sur l'ouverture de la donnée. [...] L'*open data*, excellent, maintenant qu'est-ce qu'on en fait stratégiquement ? Vers quoi on veut aller ? Qu'est-ce que ça sert ? **La problématique, ce n'est plus l'*open data*, c'est la *data***².

Cette réflexion de Léa symbolise l'évolution de la catégorisation de la donnée. Alors qu'initialement, on ne pensait que sa diffusion vers des acteurs externes, les premiers bilans des politiques d'*open data* conduisent à interroger la donnée, sa production, sa circulation et ses utilisations. Ce changement de perception se manifeste dans la création de ce poste d'administrateur général des données dont le rôle est précisément de piloter l'ensemble de la chaîne de la donnée.

*

* *

Ainsi, l'institutionnalisation de la gouvernance de la donnée se manifeste par l'autonomisation de l'enjeu « donnée » des secteurs traditionnels de politique publique. La première tentative d'institutionnalisation de la gouvernance a échoué par le refus de la direction générale de consacrer une instance transversale dédiée à cet enjeu regroupant les directions « métiers » de la communauté urbaine. Prenant note de ce refus, la direction des services informatiques va alors porter de manière autonome cet enjeu en le couvrant de la bannière du « numérique ». Outre cette autonomisation, l'institutionnalisation de la gouvernance de la donnée se caractérise par la professionnalisation de cet enjeu. Même s'il est aujourd'hui réduit à un agent, la création d'un poste d'administrateur général des données marque en effet l'émergence de professionnels spécifiquement en charge de la politique de la donnée.

Ces deux processus d'autonomisation et de professionnalisation ne sont toutefois pas suffisants pour attester de l'achèvement de l'institutionnalisation de la gouvernance de la

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

donnée. Des ressources matérielles et politiques sont en effet nécessaires pour assurer à la DINSI une capacité de pilotage de cette politique publique. Les échanges avec l'administratrice générale des données du Grand Lyon démontrent pourtant que les moyens à sa disposition sont incertains et limités.

Léa : Pour mener la politique de la donnée, il faut que j'aie les ressources en face. Il y a quand même un besoin de ressources.

Enquêteur : Parce que ça n'a pas été budgétisé pour le moment ?

Léa : Je n'en ai pas le sentiment. Je ne peux pas te dire à l'heure actuelle de quel budget je dispose. Et moi, ça me gêne beaucoup. On ne peut pas avancer sur une stratégie, même consolider la gouvernance, si en face, je n'ai pas un budget. Alors, on me dit : « *tu prendras là, tu prendras là, t'inquiète pas* ». Oui mais de quel budget je dispose ? [...] Et quand je vois qu'on nous a supprimé un poste dans l'équipe [en charge de la plateforme *open data*], ça m'interroge. Parce que l'on peut afficher une ambition, on peut mettre en avant le poste de CDO et montrer qu'il y a une vraie volonté politique, mais... le pire c'est qu'on ne tienne pas la promesse. Si on affiche des ambitions et une stratégie claire, il faut qu'on la mette en œuvre. Et là, arrivent les ressources techniques et humaines. Ouais, je m'interroge certaines fois [sur les moyens disponibles]¹.

Les ressources financières disponibles sont incertaines. La politique de la donnée ne dispose pas de crédits budgétaires autonomes. L'administratrice générale des données doit recourir ponctuellement à différentes lignes budgétaires selon les projets qu'elle souhaite développer. En outre, les ressources humaines au sein de ce service sont limitées et réduites au gré des coupes budgétaires.

La capacité de pilotage d'une politique publique de la donnée repose également sur la capacité de mobilisation de ressources politiques pour assurer la légitimité des actions de la DINSI. Le soutien politique est indispensable pour mener une politique de la donnée et parvenir à enrôler les acteurs réticents. Or, les élus du Grand Lyon sont difficilement mobilisables : peu s'intéressent à la question de la donnée, même si l'administratrice générale des données peut compter sur le soutien de la vice-présidente aux nouvelles technologies.

J'ai réalisé à quel point on avait beaucoup de chance parce que s'il n'y a pas d'élus qui portent l'open data ou la data en général, c'est une catastrophe. Dès lors qu'il n'y a pas de portage politique, y'a pas d'avancées. Donc, évidemment, sans Karine Dognin-Sauze, et tout ce qu'elle porte, on ne pourrait pas mener cette politique. On la rencontre régulièrement, elle porte les licences, elle est allée voir [la secrétaire d'État chargée du numérique] Axelle Lemaire, etc.²

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

Toutefois, au delà de cette vice-présidente, le sujet de la donnée n'est pas une priorité pour les élus de l'institution. Ils s'emparent peu de cet enjeu. Cette faible mobilisation politique s'est notamment manifestée par l'absence de défense de la politique lyonnaise d'ouverture de données lors des débats parlementaires sur le sujet.

Te dire que [les élus] n'étaient peut être pas forcément très à l'aise avec ces notions là, ça s'est traduit quand la loi Macron est sortie, dans l'article 4, elle stipule : "*les données d'information du voyageur seront ouvertes et rendues accessibles par les AOT*", on a fait un amendement, on a écrit un amendement, on l'a transmis au cabinet, transmis à l'attaché parlementaire de Gérard Collomb, qui est également l'attaché parlementaire de tous nos sénateurs, d'Annie Guillemot également. Cet amendement n'a pas du tout été porté. Il n'a été pris en charge par aucun élu. Et on a senti là qu'il y avait un problème. Si on rédigeait un amendement, qui est validé par le cabinet mais qu'après aucun élu ne s'en empare pour le porter, c'est un temps incroyable, c'est un travail qui n'est pas du tout valorisé. Et donc, notre amendement n'a pas été porté. Est-ce qu'il aurait été voté ? J'en sais rien. Mais il n'apparaît pas. Donc, c'est là que l'on a pris un peu peur.¹

Comme le révèle cet exemple d'amendement non porté par les élus de l'institution, les ressources politiques de la Métropole de Lyon, et de la DINSI en particulier, sont limitées pour mener une politique publique autonome de la donnée métropolitaine. Face à ce déficit de ressources, comment la DINSI peut-elle s'insérer durablement au cœur de la chaîne de la donnée métropolitaine ? L'étude de l'instrumentation de la gouvernance de la donnée offre une première réponse à cette question, en révélant la capacité des instruments à positionner la DINSI entre les producteurs et les utilisateurs de données.

Encadré 16 : Les élus et la donnée : le difficile passage du registre technique aux discours politiques

Les élus de la Métropole de Lyon apparaissent peu dans cette thèse. Cela n'est ni un choix délibéré de ma part, ni un angle mort de mon enquête, mais le reflet de la faible intervention des élus sur ce sujet de la donnée durant le temps de mon terrain d'étude. Cette faible appétence pose la question de la mise en politique de la donnée, c'est-à-dire son passage d'un registre technique à un discours politique². Réduite à sa technicité, la donnée est prise en charge essentiellement par les « experts » de l'administration. Le politique est mobilisé uniquement quand il faut prendre une décision et adopter une posture de fermeture

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² A l'instar de l'analyse effectuée par Yannick Barthe de la mise en politique des déchets nucléaires qui sont successivement technicisés et politisés. Y. Barthe, *Le pouvoir d'indécision*, op. cit.

du débat¹ (même s'ils ne font qu'entériner des choix décidés ailleurs, ils libèrent formellement les techniciens de la responsabilité de la décision) comme cela a été mis en évidence dans le deuxième chapitre. La difficulté pour l'administration est toutefois de parvenir à intéresser les élus au sujet de la donnée : de réussir à faire émerger des enjeux politiques là où ils ne perçoivent que des considérations techniques.

Certains élus s'emparent toutefois davantage de ces sujets techniques qu'ils traduisent politiquement. Il s'agit essentiellement d'élus positionnés à la marge du champ politique pour qui se saisir de ces sujets techniques est une opportunité d'acquérir une légitimité et de se faire une place au sein de ce champ. Novice en politique lorsqu'elle est nommée en 2008 vice-présidente à l'innovation et aux nouvelles technologies, Karine Dognin-Sauze affiche explicitement cette stratégie : « *Je voulais avoir une légitimité. Cette délégation n'existait pas. Il a fallu la créer et lui trouver une place. [...] Ma road map consistait à participer à la construction d'une ville innovante et intelligente. C'était la seule façon pour que cette délégation trouve sa place dans les grands projets métropolitains du mandat de Gérard Collomb*². » Cela s'avèrera payant puisque son poids politique est croissant : en 2014, elle obtient un poste de 12^e adjointe au Maire de Lyon en charge des relations internationales et des affaires européennes et elle passe de la 28^e à la 13^e vice-présidence la Métropole de Lyon.

III. Instrumenter la gouvernance de la donnée

Après la formalisation et l'institutionnalisation de la gouvernance de la donnée, la troisième partie de ce chapitre traite de son instrumentation. L'instrumentation de l'action publique recouvre « *l'ensemble des problèmes posés par le choix et l'usage des outils (des techniques, des moyens d'opérer, des dispositifs) qui permettent de matérialiser et*

¹ On retrouve ici une des spécificités du régime d'énonciation politique pointée par Dominique Boullier ou Tristan Riffault. Dominique Boullier, « S'orienter en politique : des partis et des ennemis » dans Jacques Lolive et Olivier Soubeyran (eds.), *L'émergence des cosmopolitiques*, Paris, La Découverte, 2007, p. 100-114 ; Tristan Riffault, « Fabriquer de la politique. La production du discours politique lors d'une campagne électorale (2013-2014) », *Réseaux*, 2015, vol. 2, n° 190-191, p. 13-44.

² Source : <http://www.lejournaldesentreprises.com/editions/69/actualite/rencontre/karine-dognin-sauze-faire-bouger-les-lignes-05-11-2010-107040.php> (Consulté le 21 mai 2016)

*d'opérationnaliser l'action gouvernementale*¹ ». La mise en œuvre de la politique publique s'incarne dans des dispositifs particuliers. Comme l'affirment Charlotte Halpern et Patrick Le Galès, l'émergence d'un secteur autonome d'action publique se révèle peu probable sans instrumentation propre². En étudiant les modalités d'opérationnalisation de la stratégie de gouvernance de la donnée dans des instruments, l'objet de cette partie est ainsi de déterminer la capacité du Grand Lyon à piloter la politique de la donnée métropolitaine.

Pour Pierre Lascoumes et Patrick Le Galès, un instrument d'action publique « *constitue un dispositif à la fois technique et social qui organise des rapports sociaux spécifiques entre la puissance publique et ses destinataires en fonction des représentations et des significations dont il est porteur*³ ». L'approche qu'ils proposent évacue la neutralité des perspectives fonctionnaliste et techniques des instruments⁴ pour prendre en compte les valeurs et les modes de régulation envisagés dont ils sont porteurs. L'instrumentation est une question politique car le choix des instruments va structurer la nature des relations entre les acteurs de la politique publique des données métropolitaines. L'instrumentation oriente les relations entre acteurs et participe au processus de recomposition des modalités de gouvernement. L'étude des instruments de la gouvernance des données met en avant la reconfiguration de la chaîne de la donnée au profit de la Métropole de Lyon et de sa direction des services informatiques.

Quatre instruments de gouvernement des données sont présentés : le catalogue des données, les licences et les conventions, un laboratoire de services urbains nommés « Le Tuba » et un dispositif de labellisation de services d'informations aux voyageurs. Ces instruments sont avant tout des dispositifs de coordination des acteurs hétérogènes que l'on souhaite associer ou dissocier de la politique de la donnée. Ils (ré)organisent les relations

¹ Pierre Lascoumes et Patrick Le Galès (eds.), *Gouverner par les instruments*, Paris, Presses de Sciences Po, 2004, p. 12.

² Charlotte Halpern et Patrick Le Galès, « Pas d'action publique autonome sans instruments propres. Analyse comparée et longitudinale des politiques environnementales et urbaines de l'Union européenne », *Revue française de science politique*, 2011, vol. 61, n° 1, p. 51-78.

³ Pierre Lascoumes et Patrick Le Galès (eds.), *Gouverner par les instruments*, *op. cit.*, p. 13.

⁴ Si Lascoumes et Le Galès utilisent de manière synonyme les notions d'instrument et de dispositif, ils opèrent une distinction entre les termes d'instrument, de technique et d'outil (l'instrument est un type d'institution, la technique un dispositif opérationnalisant l'instrument, l'outil est un élément d'une technique). Je fais le choix de les utiliser de manière équivalente dans ce chapitre, en m'accordant pour pointer les valeurs et les modes de régulation dont sont porteurs ces dispositifs.

entre gouvernants et gouvernés : entre la DINSI et les producteurs de données internes ou externes à l'institution, entre le Grand Lyon et les producteurs de données, entre le Grand Lyon et les réutilisateurs, etc. Ces instruments de gouvernance de la donnée se caractérisent par leur capacité à positionner la DINSI au centre de la chaîne de la donnée, entre producteurs et utilisateurs.

A. Le catalogue : rendre gouvernables les données

Le premier instrument, indispensable à la gouvernance des données, est le catalogue. En traduisant les données hétérogènes conservées dans différents « silos » en des entités homogènes rassemblées en un même lieu, ce dispositif de connaissance fait exister les données et les rend gouvernables.

La directive européenne INSPIRE impose aux administrations publiques un catalogue de leurs données géographiques. Dans ce cadre, la communauté urbaine de Lyon a mis en place un catalogue de ses données géographiques qui est la base de l'infrastructure de diffusion des données de l'institution (cf. Chapitre 2). Avant toute diffusion de leurs données, les producteurs doivent renseigner une fiche de métadonnées respectant une norme commune de description. Outil de gestion des métadonnées, ce catalogue est un « investissement de forme¹ », qui qualifie selon des critères communs des données hétérogènes. Il impose l'explicitation de ce qui fait la donnée. Lister les données implique de définir ce qu'est ou n'est pas une donnée. Une définition stable et générique est ainsi attribuée à des entités hétérogènes. En outre, sont associés à ces entités différents attributs : un propriétaire ou un responsable, un degré de qualité, une description, des mots-clés, etc. Ainsi, en rendant visible les données au sein d'une classification commune, non seulement le catalogue facilite l'identification des données, mais il fait également exister les données de manière performative.

Ces deux processus de standardisation des métadonnées et de mise en visibilité des données font du catalogue un instrument qui permet de « rendre gouvernable² » les données

¹ L. Thévenot, « Les investissements de forme », art cit.

² Pierre Lascoumes met en avant l'importance de ces processus de traduction pour « rendre gouvernable » un enjeu de politique publique. Pierre Lascoumes, « Rendre gouvernable : de la “traduction” au “transcodage” ». Antoine COURMONT – « Politiques des données urbaines » - Thèse IEP de Paris – 2016 343

de l'institution. Cependant, ce travail de traduction des données est aujourd'hui restreint aux données diffusées en *open data*. Absentes de ce catalogue, une large part des données métropolitaines sont « invisibles » et échappent au périmètre de la gouvernance des données.

Enquêteur : Le fait d'avoir un catalogue, ça permet aussi d'avoir une meilleure connaissance des données ?

Pauline : Oui. Sauf que tu n'as pas tout. Pour la propreté par exemple, tu as quatre données sur trente. Toutes les données que tu ne veux pas mettre en diffusion, **tu ne les vois pas**¹.

A l'instar de l'ensemble des systèmes d'information analysés par Patrice Flichy², la tension entre visibilité et invisibilité des données est manifeste dans le catalogue. Il s'agit assurément d'un instrument qui permet de « rendre visible » des informations, mais, qui, dans le même temps, contribue à invisibiliser toutes les données qui ne sont pas répertoriées. Si la centralisation des données permise par le catalogue offre à la direction des systèmes d'information une prise sur ces données, la gouvernance des données est restreinte aux seules données diffusées en *open data*³.

Pour remédier à cette méconnaissance, la DINSI souhaite étendre le catalogage à l'ensemble des données de la Métropole de Lyon et aux principaux producteurs de l'agglomération. Il s'agit ainsi de passer d'une connaissance ponctuelle à un recensement systématique et exhaustif des données. Le catalogue n'est alors plus uniquement un outil d'identification des données diffusées, mais il devient également un dispositif de gestion des données devant faciliter leur usage par des acteurs internes et externes à l'institution. Cette volonté de mieux administrer les données est partagée par la délégation générale au développement urbain (DGDU). Dans une note interne rédigée en 2014, elle préconise une meilleure « *connaissance des données disponibles au sein des services du Grand Lyon, dans une optique de partage et capitalisation de l'information* »⁴. Un « *catalogue des données (ou*

L'analyse des processus de changement dans les réseaux d'action publique. » dans CURAPP (ed.), *La gouvernabilité*, PUF., Paris, 1996, .

¹ Source : Entretien avec Pauline, administratrice de données, Grand Lyon (13 octobre 2013). Nous soulignons.

² Patrice Flichy, « Rendre visible l'information. Une analyse sociotechnique du traitement des données. », *Réseaux*, février 2013, n° 178-179, p. 55-89.

³ Ce qui ne veut pas dire que ces données ne sont pas gouvernées par ailleurs, mais elle ne le sont pas par la direction des services informatiques du Grand Lyon.

⁴ Source : Note interne « Constitution d'un catalogue commun de données « Grand Lyon » » (3 janvier 2014)

« bases de données des données ») à vocation interne au Grand Lyon »¹ doit permettre de faciliter l'usage des données pour différentes politiques publiques.

Il s'agirait de recenser les différentes sources de données, de décrire leur contenu, préciser la source de manière à ce que, face à un besoin tel que la production d'un indicateur, ou le démarrage d'un projet, un agent puisse, recenser et évaluer les sources de données disponibles susceptibles de répondre à ce besoin².

Cet instrument de catalogage des données est ainsi un outil de mise en relation des producteurs, de leurs données et des potentiels utilisateurs. Il révèle également la volonté de savoir des acteurs administratifs sur les données utilisées par l'administration en vue de les gouverner. Apparaît ainsi une nouvelle gestion des données : de ressources administratives (le support de l'activité des agents), le patrimoine informationnel devient l'objet d'une politique publique autonome et transversale.

Encadré 17 : Le catalogue, un instrument au cœur des politiques de la donnée des villes américaines

Le catalogue est un instrument central des politiques de la donnée mises en place par les administrations municipales nord-américaines. Par exemple, une des priorités du *Chief Data Officer* de Philadelphie est de mettre en place un *data inventory*, contenant des informations de bases sur chaque jeu de données, afin que l'administration puisse avoir « *the whole picture* » des données à sa disposition³. A l'instar de la Métropole de Lyon, ce catalogage s'inscrit tout autant dans la continuité de la politique d'*open data*, que dans une volonté de rationaliser l'administration municipale par une meilleure connaissance de son patrimoine informationnel. « *Data sharing between departments is an equally important outcome of the process of identifying and centralizing datasets. Rather than focusing strictly*

¹ Source : Note interne « Constitution d'un catalogue commun de données « Grand Lyon » » (3 janvier 2014)

² Source : Note interne « Constitution d'un catalogue commun de données « Grand Lyon » » (3 janvier 2014)

³ Source : Document « Open Data Strategic Plan » (Octobre 2014). Disponible à cette adresse : https://cityofphiladelphia.github.io/slash-data/phl_opendata_plan.pdf (Consulté le 19 décembre 2015)

on potentially public data, open data will become a component of a broader data management strategy that includes private/internal data¹. »

B. Licences & Conventions : devenir tiers de confiance

Les conventions de mise à disposition et les licences d'utilisation des données forment le second instrument de gouvernance de la donnée métropolitaine. Les premières sont des contrats signés entre la Métropole de Lyon et des producteurs afin d'autoriser l'utilisation de leurs données par les services du Grand Lyon et/ou la diffusion de leurs données sur la plateforme *open data*. Les secondes sont les licences encadrant l'utilisation des données diffusées, dont l'élaboration a été étudiée dans le chapitre 2. Ces instruments contractuels construisent une démarche de coopération entre les multiples acteurs engagés dans la politique de la donnée. Ces engagements partenariaux permettent la coordination d'acteurs aux intérêts divers en créant une relation de confiance entre producteurs, diffuseurs et utilisateurs de données².

Les conventions de mise à disposition visent à consentir les droits d'usage et de rediffusion des données des producteurs à la communauté urbaine de Lyon. Elles décrivent précisément les données (nature, format, qualité, support de livraison, fréquence des mises à jour, etc.), stipulent les licences qui encadreront leur diffusion, indiquent les conditions financières de fourniture. Surtout, les conventions établissent les responsabilités de chacun des signataires. Le producteur garantit la pérennité de la mise à disposition de ses données, s'engage à ne pas modifier les structures des fichiers de données sans avertir préalablement la

¹ Source : Document « Open Data Strategic Plan » (Octobre 2014). Disponible à cette adresse : https://cityofphiladelphia.github.io/slash-data/phl_opendata_plan.pdf (Consulté le 19 décembre 2015)

² Jean-Pierre Gaudin a longuement étudié la contractualisation comme modalité de mise en œuvre des politiques publiques. Selon lui, le recours à cet instrument marque la transformation vers une action publique partenariale. « Ces démarches, sous des intitulés divers, chartes, conventions, contrats pactes, mettent en forme, par la négociation d'objectifs ponctuels et de moyens ciblés, des coopérations entre des mondes multiples, à la fois partenaires et concurrents : régions, départements, communes, État, associations, entreprises » (Gouverner par contrat, p. 7). Il met en avant trois caractéristiques de ces procédures contractuelles : « 1) Un temps de discussion explicite sur les objectifs recherchés et sur les moyens correspondants ; 2) des engagements réciproques sur un calendrier d'actions et de réalisations à moyen terme ; 3) des clés de contributions (financières ou autres) conjointes à la réalisation des objectifs » (Dictionnaire des politiques publiques).

Jean-Pierre Gaudin, *Gouverner par contrat*, Paris, Presses de Sciences Po, 2007, 280 p ; Jean-Pierre Gaudin, « Contrats » dans Laurie Boussaguet, Sophie Jacquot et Pauline Ravinet (eds.), *Dictionnaire des politiques publiques*, Paris, Presses de Sciences Po, 2010, p. 164-171.

communauté urbaine. Il n'est toutefois pas responsable en cas d'erreur dans les jeux de données ou d'utilisation des données. La communauté urbaine endosse toutes les responsabilités en cas de réclamations ou de procédures impliquant la propriété ou l'usage des données du producteur. Des conventions ont aujourd'hui été signées entre la Métropole de Lyon et plusieurs acteurs publics ou privés : le SYTRAL, les opérateurs de parking (LPA, Effia, Q-Park), RhôneExpress, les Aéroports de Lyon et Bluely (groupe Bolloré).

Les licences sont des contrats signés entre le diffuseur et le réutilisateur qui cadrent l'usage des données¹. Ces licences sont un outil central du gouvernement des données puisqu'elles délimitent ce que l'on peut en faire et sous quelles conditions. Le Grand Lyon a fait le choix de mettre à disposition ses données sous trois types de licences : la licence ouverte, la licence associée et la licence engagée. Les deux dernières licences requièrent l'authentification de l'utilisateur, ainsi qu'une description de l'usage qu'il souhaite faire de ces données. Elles ont été conçues afin que la Métropole puisse « *s'assurer que les réutilisations soient compatibles avec les politiques publiques ou ne nuisent pas à des services publics* »². L'enjeu de ces licences est ainsi d'associer la donnée aux acteurs qui ne contreviennent pas aux politiques publiques, et d'empêcher les utilisations jugées contraires à l'intérêt général. Elles instaurent en creux des divisions légitimes entre les utilisateurs, en traçant une frontière entre les usages acceptables et inacceptables. Les réutilisations nuisant aux services publics ou allant à l'encontre des politiques publiques sont jugées indésirables. Les acteurs ne respectant pas ces conditions sont alors dissociés de la politique d'ouverture des données. « *L'authentification permet concrètement de suspendre l'alimentation uniquement aux réutilisateurs ne respectant pas ces obligations de qualité et de respect de l'intérêt général, sans pénaliser les autres, ce que ne permet pas l'open data* »³. Les licences sont un instrument de gouvernement des données ouvertes, qui permettent à l'acteur public de maîtriser l'usage qui sera fait de ses données⁴.

¹ J'ai détaillé dans le deuxième chapitre le processus de construction de ces licences qui stabilisent la problématisation de la politique d'ouverture des données.

² Source : Note interne, Présentation des principes de fourniture des données au Grand Lyon (17 juin 2013)

³ Source : Note interne, Présentation des principes de fourniture des données au Grand Lyon (17 juin 2013)

⁴ Une autre fonctionnalité de ces licences est qu'elles ouvrent des scènes de négociation entre acteurs. Comme nous l'indique Léa : « *La licence a prouvé son intérêt, parce qu'elle a permis la discussion avec Google* ». Antoine COURMONT – « *Politiques des données urbaines* » - Thèse IEP de Paris – 2016

Licences et conventions assurent également la pérennité de la relation entre les différents acteurs. Les producteurs s'engagent à fournir les données et la Métropole de Lyon à les mettre à disposition de manière pérenne, point crucial pour que des utilisateurs puissent utiliser de manière fiable les données. Cet aspect est particulièrement important au point que certains acteurs se demandaient s'il ne serait pas judicieux de mettre en place des conventions avec les producteurs internes de données.

Dimitri : est-ce que ce serait pertinent ou pas de conventionner également la donnée interne du Grand Lyon pour la mise à disposition sur Smart Data ?

François : C'est une bonne question. Je rebondis sur ce que tu dis, très important parce que, je prends un exemple, tu te souviens des silos à verre. La direction de la propreté avait dit on ouvre les silos à verre, deux mois après, on ferme les silos à verre. C'est vachement merdique parce qu'il y avait des utilisateurs derrière qui commençaient à bosser sur les silos à verre, et qui disent : « *ils sont plus là les silos à verre* ». Et nous, notamment dans le cadre de la volonté de développement économique autour de SmartData, on a eu encore quelqu'un, qui disait : « *c'est super fondamental que quand vous mettez une donnée, ce soit pérenne, parce que nous, on développe des services, on investit, si vous enlevez la donnée trois jours après ou si vous modifiez, c'est la merde* »¹.

Les conventions ne seront pas étendues aux producteurs internes de données, toutefois, cette discussion révèle l'importance de ces outils contractuels pour assurer durablement les relations créées entre producteurs, diffuseurs et réutilisateurs de données, qu'ils soient internes ou externes à la Métropole de Lyon. Ce lien contractuel garantit la pérennité de la relation technique entre le diffuseur et le réutilisateur. Par l'intermédiaire des licences, le Grand Lyon s'engage à diffuser les données de manière stable et continue, à informer les réutilisateurs en cas de dysfonctionnement et à y remédier dans un « délai raisonnable² ». Le fait de suivre une règle stabilise la relation en créant des situations d'irréversibilité. En signant la convention, le producteur s'engage à continuer à mettre à disposition sa donnée et à ne pas revenir à une situation antérieure.

Les licences et les conventions sont ainsi des instruments contractuels visant à formaliser des démarches partenariales d'action publique. Ils sont également des outils participant à la création d'une relation de confiance entre les acteurs dans une situation initiale

Comme nous l'indique Léa, les licences ont permis d'instaurer un dialogue avec certaines entreprises réutilisatrices sur lesquelles le Grand Lyon n'a pas d'autres « prises ».

¹ Source : Journal de terrain, réunion Gouvernance de la donnée (14 octobre 2013)

² Source : Document « Licence engagée » de la Métropole de Lyon. Disponible à cette adresse : <http://data.grandlyon.com/files/2015/03/LicenceEngagee.pdf> (Consulté le 19 décembre 2015)

d'incertitude. En effet, en formalisant les relations entre les différents acteurs de la chaîne de la donnée, les licences et les conventions réduisent les ambiguïtés et les incertitudes en stipulant les engagements et en attribuant les responsabilités de chacun des acteurs. A ce titre, ces contrats créent des situations de confiance nécessaires à l'établissement de relations entre acteurs aux intérêts hétérogènes. La responsabilité de l'usage des données est en effet une crainte majeure qui peut nuire à la coordination des acteurs. Par exemple, lors d'une réunion préparatoire, les personnes en charge des données de l'aéroport de Lyon s'interrogeaient : « *Si les personnes qui utilisent les données ne garantissent pas les mises à jour et qu'un client rate son avion, qui est responsable ?* »¹. La signature d'une convention les déchargeant de toute responsabilité a contribué à la mise à disposition de leurs données sur la plateforme de la Métropole de Lyon. De même, les licences avec authentification créent de la confiance chez les producteurs et facilitent la mise à disposition de leurs données.

Je pense que l'on a rassuré en fait. On a rassuré parce qu'avec notre stratégie de mise à disposition, on disait : « *on ne met pas à disposition de n'importe qui, si vous voulez vous pouvez savoir à qui vous mettez à disposition, vous pouvez voir les services qu'il fait. Si vous n'êtes pas d'accord, vous nous le dites.* » Et ça, ça a quand même pas mal rassuré les gens. [...] Le fait d'adresser à la fois les gens en situation de monopole plus les réutilisations on va dire fantaisistes, ça a plutôt rassuré les gens, et c'est ce qui fait que ça a marché².

Licences et conventions sont des instruments de coordination des acteurs de la chaîne des données ouvertes. Comme l'indique cet interlocuteur, les licences ne sont pas uniquement un outil pour encadrer les usages des données, mais également un dispositif visant à encourager la mise à disposition des données. « *Ces licences, c'est un vrai levier pour l'ouverture. Elles créent un cadre de confiance pour nos partenaires*³ ». En réduisant l'incertitude, les licences produisent de la confiance et facilitent l'association des producteurs à la politique d'*open data* du Grand Lyon. Soumises à l'évolution du cadre législatif national, ces licences avec authentification pourraient être supprimées, ce qui fait craindre une rupture des relations entre le Grand Lyon et certains partenaires. « *La loi Valter qui transpose la directive PSI, ça met vraiment en danger [les licences], et du coup ça met en danger nos relations avec nos partenaires. On doit signer avec Toshiba une convention pour publier les*

¹ Source : Journal de terrain, Echanges avec Aéroports de Lyon (24 mars 2014).

² Source : Entretien avec Harald, chargé de mission mobilité urbaine, Grand Lyon (19 janvier 2015)

³ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

données énergie, pour eux, ils veulent les publier mais uniquement si c'est en licence engagée voir associée, on va être en grosse difficulté si on ne peut les ouvrir qu'en open data¹ ». Les licences sont ainsi des instruments favorisant les relations entre producteurs, diffuseurs et utilisateurs de données. En réduisant les incertitudes, ils favorisent l'établissement de partenariats entre ces acteurs aux intérêts hétérogènes.

Les conventions et les licences sont des instruments de gouvernement de la donnée qui assurent la mise en relation d'acteurs divers en créant de la confiance². La confiance est indissociable de l'incertitude suscitée par l'*open data*. C'est parce que les producteurs s'interrogent sur l'usage qui sera fait de leurs données (« *N'ira-t-il pas à l'encontre de la politique publique ?* »), ou que les utilisateurs se demandent si la fourniture des données sera pérenne que la question de la confiance est soulevée. Le contrat est producteur de confiance par la formalisation de la relation qu'il induit. Il objective le rapport entre les acteurs de la chaîne de la donnée, garantit les identités de chacun des acteurs et instaure les engagements de chacun. En établissant des règles, les conventions et les licences permettent à chaque participant de la relation d'anticiper les comportements des acteurs et ainsi de réduire l'incertitude et de créer une situation de confiance. En signant conventions et licences, la Métropole de Lyon se porte garant dans l'interaction entre producteurs et utilisateurs : elle se positionne comme un tiers de confiance entre ces deux parties.

C. Le label : gouverner par les modèles

Le troisième instrument de gouvernement des données est le dispositif de labellisation de services d'information aux voyageurs³. Cet instrument, mis en place par la Direction de la

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² Sur l'apparition du problème de la confiance dans les situations de désobjectivation des rapports sociaux, voir Anne Girard et Violaine Roussel, « Une question de confiance. », *Raisons Pratiques*, 2003, vol. 2, n° 10, p. 171-186.

³ Le label comme instrument d'action publique a été analysé par Bergeron et al. dans le cadre des politiques de l'obésité et de la consommation durable, et par Renaud Epstein pour analyser le « gouvernement à distance » des villes par l'Etat.

Henri Bergeron, Patrick Castel et Sophie Dubuisson-Quellier, « Gouverner par les labels. Une comparaison des politiques de l'obésité et de la consommation durable », *Gouvernement et action publique*, 2014, vol. 3, n° 3, p. 7-31 ; R. Epstein, « La gouvernance territoriale : une affaire d'État La dimension verticale de la construction de l'action collective dans les territoires. », art cit.

voirie de la communauté urbaine, permet aux entreprises réutilisatrices des données mobilité d'utiliser la marque « OnlyMoov » qui leur offre « *une notoriété et un soutien notable à la diffusion de leurs services* »¹. Pour obtenir ce label, ces services doivent respecter la charte « Information Mobilité Durable » élaborée par le Grand Lyon.

Les services répondant à la charte devront en particulier comprendre a minima les modes et services suivants : services des Transports en commun lyonnais (TCL), services Vélov, vélo. Les services suivants pourront être intégrés suivant la disponibilité des données : Train express régional (TER), autopartage, covoiturage. L'information fournie doit être fiable et lorsque l'information routière est comprise dans le service, elle devra utiliser les données historiques de trafic permettant d'intégrer les congestions récurrentes du trafic routier aux temps de parcours en voiture et présenter le coût d'usage de la voiture avec les temps de parcours².

Le Grand Lyon a défini précisément un ensemble de critères qui traduisent les grands principes de sa politique de mobilité urbaine (nombre minimal de modes de transport intégré, fiabilité de l'information fournie, neutralité du service, etc.). Après examen des dossiers de candidature, le Grand Lyon accorde un droit d'usage de sa marque « OnlyMoov » à l'entreprise et met en avant ses services d'information dans sa communication institutionnelle. Le label est ainsi conçu comme un signe distinctif valorisant les applications respectant les critères définis par le Grand Lyon. L'entreprise retire de cette reconnaissance une visibilité supplémentaire sur le marché des services d'information aux voyageurs.

Pour le Grand Lyon, ce dispositif de labellisation permet d'influer sur des entreprises autonomes en les incitant à développer des services respectant les principes de sa politique publique³. En distinguant quelques services, il met en avant les « bonnes pratiques » de réutilisation des données ouvertes et les érige en modèle pour les autres réutilisateurs. Ce « gouvernement par les modèles⁴ » offre à la Métropole de Lyon une capacité de pilotage de

¹ Source : Décision n° B-2014-0330 du bureau du 8 septembre 2014 de la communauté urbaine de Lyon.

² Source : Décision n° B-2014-0330 du bureau du 8 septembre 2014 de la communauté urbaine de Lyon.

³ La mission gouvernementale Etalab a mis en place un dispositif de labellisation qui vise les producteurs plutôt que les réutilisateurs de données. Le label qualité « référentiel national » est accordé au producteur qui « s'engage à fournir la donnée brute, à respecter strictement son cycle de mise à jour, et à répondre à J+2 à tout signalement ou demande ». En retour, son jeu de données est mis en avant sur la plateforme data.gouv.fr et est mis à disposition sous forme d'API.

<https://www.data.gouv.fr/fr/faq/producer> (Consulté le 29 novembre 2015)

⁴ Vincent Béal, Renaud Epstein et Gilles Pinson, « La circulation croisée. Modèles, labels et bonnes pratiques dans les rapports centre-périphérie », *Gouvernement et action publique*, 2015, vol. 3, n° 3, p. 103-127.

l'utilisation des données, sans contraindre les acteurs qui conservent leur autonomie. Toutefois, pour que cet instrument remplisse son rôle, deux conditions sont nécessaires. Il faut d'une part qu'il y ait un marché suffisamment concurrentiel pour que le label soit bénéfique à l'adoption de ce service. D'autre part, il est nécessaire que les acteurs de ce marché aient le désir de se distinguer¹. Ces deux conditions ne semblent pas réunies dans le cas la politique *d'open data* du Grand Lyon². Un an après l'entrée en vigueur de ce dispositif de labellisation, un seul acteur a candidaté à ce label : l'entreprise Cityway qui a développé l'application Optymod... en partenariat avec le Grand Lyon !

D. *Le living lab : mettre en relation producteurs et utilisateurs*

Le dernier instrument de la gouvernance des données métropolitaines est « Tuba », un *living lab* créé à l'initiative de la Métropole de Lyon en vue d'accompagner les entreprises utilisatrices de données dans le développement de nouveaux services urbains.

Encouragé par l'Union européenne³, le *living lab* lyonnais trouve son origine dans la volonté du pôle de compétitivité Imaginove de réaliser un « *laboratoire d'usages [...]* *proposant l'observation des comportements numériques des utilisateurs et consommateurs* »⁴. Ce « *laboratoire vivant* » vise à étudier les usages numériques des consommateurs afin de « *construire et développer des services à valeur ajoutée pour l'usager* »⁵. Ce projet est ensuite repris par la communauté urbaine de Lyon qui l'adosse à sa démarche de ville intelligente et à sa politique *d'open data*. Le « *living lab* » devient « *une plateforme d'innovation et d'expérimentation, visant à créer de nouveaux services urbains liés au numérique* »⁶. Il

¹ H. Bergeron, P. Castel et S. Dubuisson-Quellier, « Gouverner par les labels. Une comparaison des politiques de l'obésité et de la consommation durable », art cit.

² L'entreprise Keolis, en charge du réseau de transport en commun rennais, a mis en place un dispositif similaire de labellisation, qui rencontre davantage de succès puisque 9 applications sont aujourd'hui labellisées.

<https://data.keolis-rennes.com/fr/le-projet/labellisation.html> (Consulté le 29 novembre 2015)

³ L'Union européenne a mis en place une stratégie de développement des *living labs* dans le but de favoriser l'innovation et l'expérimentation.

⁴ Source : Communiqué de presse d'Imaginove publié suite à son assemblée générale du 27 juin 2011.

⁵ Source : Présentation d'Étienne Guerry, chef de projet Living Lab, Imaginove (25 novembre 2011). Disponible à cette adresse : <http://www.alcotra-innovation.eu/newseventi/doc/CR25nov11.doc.pdf> (Consulté le 19 décembre 2015)

⁶ Source : Journal de terrain, réunion avec Lucie, chef de projet, Grand Lyon (6 décembre 2012)

s'inscrit dans le prolongement de la politique d'*open data*, puisqu'il est en charge de l'animation de la démarche afin de favoriser les réutilisations des données ouvertes.

Pour nous, l'*open data* n'est pas dissociée du Living lab que nous sommes en train de monter. Le Grand Lyon crée une entité qu'il pilote et dans laquelle des entreprises vont pouvoir disposer d'un terrain de jeu pour créer des services, co-construire des services avec des usagers sur la base de données que nous mettons à leur disposition et de données qu'elles-mêmes vont pouvoir récupérer. Il s'agit de créer une sorte de super *open data* public / privé pour travailler sur des usages avec des citoyens. C'est un vrai laboratoire des usages, pour créer de nouveaux services. Il y a des start-up qui ont des idées extraordinaires ; elles ont juste besoin d'un lieu et de données pour expérimenter¹.

Le « living lab » est défini comme un instrument d'accompagnement des entreprises dans l'appropriation des données urbaines mises à disposition par le Grand Lyon en vue de la création de nouveaux services à destination des « citoyens ». Une certaine conception de l'*open data* est inscrite dans cet instrument. La figure des réutilisateurs est celle des petites entreprises « innovantes », créatrices de « services » numériques, regroupées sous le terme de « start-up ». Si l'on souhaite les associer à ce processus d'innovation, les « citoyens » sont avant tout définis comme des consommateurs de services commercialisés par des entreprises privées. Enfin, les grandes entreprises sont appréhendées comme financeurs, fournisseurs de données et à la recherche d'idées et d'innovations fournies par des acteurs externes.

Ce projet a donné naissance au « Tuba ». Le « Tuba » se définit comme un « *laboratoire urbain collaboratif* »². Les trois termes de cette définition sont importants puisqu'ils soulignent les spécificités de cet instrument. « Laboratoire » fait référence aux méthodes employées qui s'inspirent des démarches scientifiques expérimentales. « Collaboratif » met en avant le rôle d'intermédiaire entre différents acteurs qu'il entend jouer. « Urbain » souligne que la ville est son « terrain d'expérimentation » afin « *d'atteindre son objectif de création de la ville intelligente* »³.

Cet instrument se caractérise en premier lieu par son rôle de mise en relation de différents acteurs. Par l'intermédiaire des données, le Tuba entend rapprocher des « grandes

¹ Source : Interview de Philippe Vial, directeur des systèmes d'information, Millénaire3, 30 novembre 2012, <http://www.millenaire3.com/interview/services-urbains-et-donnees-publiques> (Consulté le 15 décembre 2015)

² Source : <http://www.tuba-lyon.com/tuba-accueil/tuba-cest-quoi> (Consulté le 21 novembre 2015)

³ Source : <http://www.tuba-lyon.com/tuba-accueil/tuba-cest-quoi> (Consulté le 21 novembre 2015)

entreprises », des « porteurs de projets (« startup ») », des « pôles de compétitivité, laboratoires de recherches et des clusters », et des « citoyens »¹. Etrangement, les acteurs publics sont absents de cette présentation alors qu'ils sont à l'origine de ce projet et qu'ils participent à son financement². Le Tuba est porté par l'association Lyon Urban Data dont sont membres des collectivités territoriales (Métropole de Lyon, Région Rhône-Alpes), des entreprises (EDF, ERDF, Keolis, SFR, Sopra, Veolia), des TPE et PME (ForCity, Esri France), des pôles de compétitivité et des organismes de recherche et d'enseignement supérieur (Imaginove, LUTB, Cluster Edit, INSA, Liris, IMU, etc.).



¹ Source : <http://www.tuba-lyon.com/tuba-accueil/tuba-cest-quoi> (Consulté le 21 novembre 2015)

² Le budget 2014 de l'association est de 506 850 euros provenant pour moitié de subventions de la communauté urbaine de Lyon (140 850€) et de la Région Rhône Alpes (114 000€) et pour moitié des cotisations de cinq entreprises membres à hauteur de 50 000€ chacune (EDF, Veolia Environnement, SFR, SOPRA, Kéolis). En 2015, l'entreprise ERDF devient membre de l'association pour un montant de cotisation identique, tandis que l'Europe s'ajoute aux subventionnaires publics (71 330€ de subventions pour chaque acteur public). En 2013, la communauté urbaine de Lyon (68 000€) et EDF, Veolia Environnement et SFR (20 000€ chacune) avaient participé au financement du projet de living lab.

Source : Délibération n°2013-4087 du Conseil de communauté du 26 septembre 2013 ; Délibération n°2014-0212 du Conseil de communauté du 10 juillet 2014 ; Délibération n°2015-0428 du Conseil métropolitain du 6 juillet 2015

Illustration 31 : Visuel de présentation des « contributeurs » du Tuba¹

La seconde spécificité de cet instrument est l'application de méthodes, issues de la recherche, du design et des incubateurs de technologies numériques, à l'action publique et la création de services urbains. Tuba est un « laboratoire », à ce titre, il utilise des méthodologies expérimentales², c'est-à-dire tester des services à petite échelle, dans l'espace urbain, avant de les déployer³. Ces méthodes sont complémentaires aux démarches de design de prototypage rapide, où les utilisateurs sont associés rapidement à l'élaboration du produit. Dans le cadre du Tuba, les « citoyens » sont perçus comme « *un maillon indispensable au processus de création* », ils doivent « *s'impliquer dans le processus d'innovation* » en « *testant la pertinence de ces nouveaux services* »⁴. Enfin, le Tuba répond à une logique d'innovation ouverte, fortement développée dans l'économie numérique, qui vise, pour une entreprise (ou une collectivité), à trouver des connaissances, des méthodes ou des technologies à l'extérieur.

Par son rôle de mise en relation d'acteurs divers et la promotion de méthodes expérimentales, le « Tuba » est un instrument de la politique des données de la Métropole de Lyon. Il doit permettre à la collectivité d'associer à l'*open data* des utilisateurs dont le profil et les usages des données correspondent à sa politique publique de Métropole intelligente. L'imaginaire associé à cette dernière, et véhiculé par le Tuba, est en effet celui de jeunes entreprises développant, sur le territoire, des services innovants, à partir de données publiques ou privées, à destination des citoyens. Le Tuba prolonge la politique de développement économique de la Métropole de Lyon en lui permettant de répondre aux besoins des entreprises utilisatrices de données.

Léa : Dans notre rôle d'accompagnement du développement économique, [on doit] aussi travailler avec les startups et les entreprises qui travaillent avec nos données pour leur demander de quoi elles ont besoin. Ça ça me semble très important. Et le Tuba, pour moi, ça c'est leur job, mais il faut

¹ Source : <http://www.tuba-lyon.com/tuba-accueil/tuba-cest-quoi> (Consulté le 21 novembre 2015)

² Le profil de son personnel reflète cette expertise méthodologique : titulaire d'un doctorat en chimie physique, sa directrice travaillait au sein du département R&D d'un groupe industriel et ses deux chargés de mission ont une formation en management des technologies de l'innovation et en design industriel.

³ Pour une analyse d'un laboratoire d'innovation dans le domaine des politiques de l'éducation, voir Ben Williamson, *Testing governance: the laboratory lives and methods of policy innovation labs.*, [Working paper], 2015.

⁴ Source : <http://www.tuba-lyon.com/tuba-accueil/tuba-cest-quoi> (Consulté le 21 novembre 2015)

qu'on soit à leur écoute aussi. Il faut que l'on fasse remonter, s'ils nous disent voilà, on a besoin de telle donnée, il faut que l'on soit en capacité d'aller les chercher.

Enquêteur : le Tuba est aussi utilisé pour aller chercher des données en interne ?

Léa : Oui. Ça fait levier. On leur dit : « vos données elles seront utilisées dans tel cadre, allez on y va ! ». Par exemple, il y a trois challenges pour la Cop 21, il y a une donnée qui était nécessaire, on va l'ouvrir, on a réussi¹.

Comme le souligne Léa, le Tuba joue un double rôle. D'une part, il permet à la Métropole de Lyon d'accompagner les entreprises dans leurs processus d'innovation en mettant à disposition les données dont elles ont besoin. D'autre part, en rendant concret l'usage des données, il est un dispositif d'intéressement visant à inciter les producteurs à diffuser leurs données. Ainsi, le Tuba semble être pour la Métropole de Lyon un instrument central de la gouvernance des données métropolitaines par sa capacité à mettre en relation producteurs et utilisateurs de données. Toutefois, un an après sa création, il est difficile de tirer un premier bilan des réalisations concrètes du Tuba au regard de ces discours initiaux. Remplit-il véritablement son rôle d'instrument de la politique métropolitaine de la donnée ? Son rapprochement d'autres institutions publiques et privées ne traduit-il pas au contraire une prise de distance vis-à-vis de la Métropole de Lyon ?

*

* *

Le catalogue, les licences et conventions, le « Tuba » et le label sont aujourd'hui les quatre instruments principaux du gouvernement de la donnée métropolitaine. Le catalogue de données est un outil de savoir qui permet de représenter et de connaître les entités à gouverner. Les licences et conventions sont des dispositifs contractuels qui réduisent l'incertitude et encadrent les actions de chacun des acteurs. Le « Tuba » est un laboratoire favorisant la mise en relation des producteurs et des utilisateurs. Enfin, le label incite les utilisateurs à suivre les principes d'utilisation des données définis par la collectivité. Ces instruments sont autant de médiations qui viennent complexifier la chaîne de la donnée et qui permettent à la Métropole de Lyon – et à sa direction des services informatiques – de réguler les relations entre producteurs et utilisateurs de données.

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

Ces instruments de gouvernement de la donnée sont essentiellement conventionnels et incitatifs¹. Tout en reconnaissant l'autonomie de chaque acteur de la chaîne de la donnée, la Métropole de Lyon vise par ces instruments à assurer leur coordination et ainsi à piloter la gouvernance de la donnée métropolitaine. Ce mode de coopération incitatif caractérise la politique de la donnée. L'action de la Métropole s'appuie sur un répertoire de la régulation : on organise les relations entre acteurs, on coordonne des intérêts variés, tout en laissant une large marge de manœuvre aux différents acteurs pour agir. La politique publique de la donnée est ainsi une « action publique procédurale » au sein de laquelle « *le processuel l'emporte sur le substantiel*² » pour reprendre les mots de Pierre Lascoumes au sujet de l'environnement. La Métropole ne fixe pas réellement de contenus et d'objectifs à atteindre, mais elle met en place des procédures pour cadrer les modalités d'interaction entre les différents acteurs concernés.

Les instruments permettent d'établir des associations fortes entre données, producteurs, utilisateurs et la Métropole de Lyon. Ils visent à rendre irréversibles les interactions entre acteurs en les inscrivant dans des boîtes noires³. Le catalogue stabilise l'identité des données, préalable nécessaire à leur gouvernement. Les licences créent de l'irréversibilité en formalisant les relations entre le Grand Lyon et les producteurs ou les utilisateurs. Avec le Tuba, la Métropole s'allie des entreprises utilisatrices de données et créatrices de services. Avec les labels, elle s'assure que ces services n'aillent pas à l'encontre d'associations préexistantes avec les acteurs en charge des politiques publiques métropolitaines. C'est dans sa capacité à « s'asseoir sur des boîtes noires⁴ » pour solidifier le réseau d'acteurs qu'elle a constitué que la Métropole de Lyon tire sa légitimité à gouverner la donnée.

¹ Pierre Lascoumes et Patrick Le Galès (eds.), *Gouverner par les instruments*, op. cit., p. 357-370.

² Pierre Lascoumes, *L'éco-pouvoir. Environnements et politiques.*, Paris, La Découverte, 1994, p. 104.

³ Proposé par la sociologie de l'acteur-réseau, le concept de boîte noire fait référence à ce sur quoi il n'est plus nécessaire de revenir.

⁴ Michel Callon et Bruno Latour, « Le grand Léviathan s'apprivoise-t-il ? » dans Madeleine Akrich, Michel Callon et Bruno Latour (eds.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'Ecole des Mines, 2006, p. 11-32.

IV. Légitimer la gouvernance de la donnée

Après avoir analysé la formalisation, l'institutionnalisation et l'instrumentation de la politique publique de la donnée métropolitaine, la dernière partie de ce chapitre s'emploie à étudier sa légitimité. Par l'intermédiaire de sa direction des services informatiques, la Métropole de Lyon souhaite gouverner les données du territoire et agir sur des acteurs non soumis à son pouvoir souverain¹. Quel est alors le bien-fondé de son intervention sur les données métropolitaines ? Sur quoi repose la légitimité de la politique publique de la donnée de la Métropole de Lyon ?

La science politique distingue deux formes principales de légitimité de l'action publique en se basant sur l'opposition entre gouvernement *par* et *pour* le peuple². La première repose sur les *inputs* de la politique publique : cette dernière est légitime parce qu'elle est le reflet de la représentation démocratique. La seconde réside dans ses *outputs* : l'action publique est légitime parce que ses résultats favorisent le bien-être collectif. Dans notre cas, le premier modèle d'une légitimité d'un gouvernement *par* le peuple pose problème puisque la donnée n'est pas un enjeu politique, dans le sens où il n'est pas débattu dans les arènes démocratiques, et ne constitue pas, pour les électeurs, un critère de détermination de leur choix. En outre, même si cela devenait le cas, la question continuerait à se poser dans la mesure où la Métropole de Lyon souhaite agir sur les données d'autres institutions agissant sur son territoire. Le second modèle de légitimité d'un gouvernement *pour* le peuple nécessite de fournir une évaluation quantifiable de cette action publique de la donnée³. Or, aujourd'hui, la Métropole de Lyon n'est pas en mesure d'estimer les bénéfices de la politique de la donnée

¹ L'idéal-type wébérien de la domination légale rationnelle, basée sur la bureaucratie et le droit, trouve ses limites dans la mesure où la Métropole de Lyon ne dispose pas d'une légitimité légale pour agir sur ce domaine de politique publique, et, en outre, n'est pas autonome pour mener cette action publique, puisqu'elle est dépendante d'autres acteurs (producteurs, utilisateurs), sur lesquels elle doit s'appuyer pour assurer le succès de cette politique publique.

Sur les limites du modèle wébérien face aux nouvelles formes d'activités gouvernementales, voir Patrice Duran, « Légitimité, droit et action publique », *L'Année sociologique*, 2009, vol. 59, n° 2, p. 303-344.

² Pascale Laborier, « Légitimité » dans Laurie Boussaguet, Sophie Jacquot et Pauline Ravinet (eds.), *Dictionnaire des politiques publiques*, 3e édition., Paris, Presses de Sciences Po, 2010, p. 335-343.

³ Pour une discussion critique de ces principes de légitimité de l'action publique, voir Patrice Duran, « Action publique et pouvoir politique. La difficile conciliation de la légitimité et de l'efficacité » dans *Penser l'action publique*, Paris, LGDJ, 1999, p. 75-112.

qu'elle commence à mettre en œuvre. L'évaluation de cette politique publique processuelle est d'autant plus compliqué qu'elle ne se fixe pas d'objectifs clairs à atteindre. Dès lors plutôt que de rechercher les principes de la légitimité du Grand Lyon à gouverner la donnée métropolitaine, il est préférable d'étudier les modalités concrètes de légitimation de l'action publique¹, c'est-à-dire l'ensemble des actions par lesquelles la Métropole de Lyon tend à établir sa légitimité sur le gouvernement de la donnée métropolitaine.

Dans cette perspective, l'objet de la démonstration est d'analyser comment la Métropole de Lyon, et sa direction des services informatiques, acquièrent la légitimité pour gouverner les données produites et utilisées par d'autres acteurs. Dans la perspective de la sociologie de l'acteur-réseau², la légitimité du Grand Lyon repose sur sa capacité à construire des réseaux d'acteurs stables, à connecter des mondes sociaux et à agir en traducteur des intérêts de chacun pour se constituer comme tiers de confiance de la chaîne de la donnée métropolitaine. Afin de se voir reconnaître cette autorité, la Métropole de Lyon s'appuie sur des médiations qui solidifient le réseau créé en associant durablement l'ensemble des acteurs.

La légitimation s'est déjà exprimée au cours des trois processus précédemment présentés : en mettant en forme la donnée comme un enjeu à gouverner, en élaborant des institutions en charge de cet enjeu et en développant des instruments pour mettre en œuvre cette politique publique, la Métropole de Lyon a d'ores et déjà œuvré au processus de traduction qui doit asseoir son pouvoir sur la donnée métropolitaine. A partir du cas de la mobilité, je reviens dans une première partie sur ce processus de légitimation de la Métropole de Lyon par la construction d'un réseau d'acteurs. La solidité de ce réseau est toutefois mise à l'épreuve, ce que je présenterai dans un second temps, rendant incertaine la légitimité du Grand Lyon à gouverner la donnée métropolitaine.

¹ Jacques Lagroye définit la légitimation comme « *un ensemble de processus qui rendent l'existence d'un pouvoir coercitif spécialisé tolérable sinon désirable, c'est-à-dire qui le font concevoir comme une nécessité sociale voire comme un bienfait* ». Jacques Lagroye, « La légitimation » dans Madeleine Grawitz et Jean Leca (eds.), *Traité de science politique*, Paris, Presses Universitaires de France, 1985, p. 395-467.

² M. Callon et B. Latour, « Le grand Léviathan s'apprivoise-t-il ? », art cit.

A. Une position de « tiers de confiance » : l'exemple de la mobilité

La mobilité urbaine permet de pointer les recompositions politiques liées à la mise en œuvre d'une gouvernance de la donnée métropolitaine. L'analyse de la mise en place d'une Centrale de mobilité par la communauté urbaine de Lyon met en lumière les enjeux de légitimation de cette institution pour être en mesure de gouverner la donnée de ce secteur d'action publique. Pour rappel, le projet de centrale de mobilité consiste à rassembler les données de l'ensemble des opérateurs de transport de l'agglomération puis à les mettre à disposition d'entreprises afin qu'elles développent des services d'information à destination des voyageurs.

La première interrogation concerne l'acteur devant mettre en œuvre cette Centrale de mobilité. Cette question n'a semble-t-il pas suscité de débat. Si l'on en croit les responsables de ce projet, le Grand Lyon s'en est emparé « naturellement ». Du fait de son indépendance vis-à-vis des réseaux de transport en commun, la communauté urbaine de Lyon s'est sentie légitime pour se positionner comme un « tiers de confiance¹ », intermédiaire entre les producteurs et les réutilisateurs de données.

Pourquoi le Grand Lyon plutôt qu'un autre acteur ? Parce qu'en fait, grosso modo, il y a deux acteurs sur l'agglomération sur l'information transport que sont le SYTRAL et le Grand Lyon. Le SYTRAL est extrêmement lié au transport en commun, voir uniquement transport en commun, et le public que l'on veut adresser, ce sont ceux qui ont un usage excessif de la voiture. Que n'adresse pas le SYTRAL donc. L'information multimodale, ça consiste à faire une information sur tous les modes, y compris la voiture, ce que le SYTRAL ne fera jamais. Donc on a repris la main là dessus en intégrant l'ensemble des modes et derrière en cohérence avec la stratégie, il était assez naturel qu'on se positionne comme agrégateur de données. D'une part pour nos propres services et d'autre part si on voulait donner accès à des données, à des tiers, de manière simple et surtout complète. [...] Donc c'est assez naturellement que nous ait venu l'idée structurante, de dire : « ***Il faut qu'il y ait un endroit où on trouve tout, et il vaut mieux que ça soit le public qui joue ce rôle de tiers de confiance entre les producteurs de données et les réutilisateurs*** ». Parce que d'abord, on a la légitimité que n'aura jamais un autre acteur. Ensuite on veut garder le contrôle quand même sur la réutilisation².

Le responsable du projet Centrale de mobilité résume le rôle de l'institution devant mener ce projet. Elle doit rassembler les données de l'agglomération, être un tiers de

¹ La notion de « tiers de confiance » est employée à la suite des acteurs interrogés. Elle ne repose pas sur une conceptualisation *a posteriori* du positionnement du Grand Lyon dans la gouvernance des données de mobilité.

² Entretien avec Harald, Chargé de mission Mobilité urbaine – Grand Lyon (19 janvier 2015)

confiance pour l'ensemble des acteurs et conserver une maîtrise de l'usage des données. Ces trois fonctions font émerger une nouvelle forme d'exercice du pouvoir, qui ne consiste plus à gouverner un réseau de transport par l'intermédiaire de données, mais à assurer l'association d'acteurs et de données hétérogènes pour gouverner la mobilité métropolitaine. Cela lui procure un statut particulier d'institution de coordination, qui n'est ni purement politique, ni spécifiquement technique, mais qui englobe les deux dimensions afin d'aligner les intérêts de l'ensemble des actants. Ce travail de coordination se joue à deux niveaux de la chaîne d'information mobilité : en amont auprès des opérateurs de transports produisant les données et en aval auprès des utilisateurs de ces données. Ce positionnement biface permet au Grand Lyon de devenir un tiers de confiance et d'acquiescer la légitimité pour gouverner la mobilité à l'échelle métropolitaine.

1) Rassembler les données

Les données relatives à la mobilité sont produites et conservées par chacun des exploitants des réseaux de transport. La première fonction assignée au Grand Lyon a été de rassembler ces données au sein d'une même infrastructure et de les articuler. Comme cela a été détaillé dans les chapitres précédents, l'accès à ces données a été obtenu à l'issue de longues négociations auprès de chacun des producteurs. Il est le résultat d'un travail de mobilisation et d'intéressement des producteurs dont l'ampleur était largement sous-estimée au début du projet de Centrale de mobilité.

Enquêteur : A posteriori, quels sont les plus gros challenges que vous avez rencontrés dans la mise en œuvre du projet ?

François : La difficulté la plus importante, c'est l'obtention des accords des fournisseurs de données à participer à une démarche alors qu'ils n'ont pas forcément en tête les enjeux, donc arriver à mobiliser des acteurs, quand ce n'est pas immédiatement dans leur stratégie, je trouve que c'est une vraie difficulté, qui a été menée par beaucoup, beaucoup de lobbying et de démarchage en fait. Harald a fait énormément de travail auprès des partenaires mobilité pour obtenir des données. Ça je trouve que c'est une vraie difficulté. [...] C'est un gros chantier d'aller trouver des partenaires, leur montrer l'intérêt. Le volet obtention de données, il est quand même compliqué.

Enquêteur : Et ça tu ne l'identifiais pas comme une difficulté en tant que telle ?

François : Pas autant. Non, non, pas autant. En fait, je pensais plus aux difficultés techniques, et en fait, c'est vraiment l'inverse qui s'est avéré. La difficulté technique elle est faible par rapport à la difficulté d'obtenir un accord¹.

Pour obtenir l'accès aux données, le Grand Lyon ne s'inscrit pas dans une posture contraignante. Il ne dispose d'aucune légitimité pour imposer à ces acteurs souverains l'ouverture de leurs données. Les accords sont obtenus à l'issue de réunions de négociation visant à intéresser le producteur à l'ouverture de ses données en lui présentant les bénéfices qu'il peut retirer à s'inscrire dans cette démarche (cf. Chapitre 3). La Centrale de mobilité est exposée comme une opportunité de désengorgement de son réseau de transport et lui permet de bénéficier de l'image « novatrice » associée à ce projet². L'échange de données est conclu par la signature d'une convention de mise à disposition définissant les obligations et les droits de chacune des parties. Par ce contrat, le producteur est lié au projet et devient un « partenaire » du Grand Lyon. La contractualisation est le premier instrument d'association des acteurs et de leurs données.

Ce long travail d'intéressement effectué, le Grand Lyon assure ensuite la coordination entre ces acteurs par l'intermédiaire de leurs données. Comme décrit précédemment (cf. Chapitre 4), il met en relation les données hétérogènes au travers d'un référentiel géographique commun. Cette double capacité à centraliser et à assembler des données participe à la légitimation du Grand Lyon comme institution de gouvernement de la mobilité urbaine. Le Grand Lyon joue un rôle d'intermédiaire qui permet la mise en relation entre les différents acteurs du transport urbain.

2) Réguler l'usage des données

La recomposition du rôle de l'institution se joue également en aval de la chaîne des données. Le Grand Lyon ne fait pas l'usage direct de ces données, mais il les met à disposition d'entreprises afin qu'elles développent des services d'informations à destination des citoyens. Cette politique d'*open data*, en changeant le modèle économique des services

¹ Entretien avec François, Chef de projet informatique – Grand Lyon (11 février 2015)

² A contrario, l'image négative est un moyen de pression pour les acteurs réticents. Ainsi, à propos d'un opérateur n'ayant pas souhaité mettre ses données à disposition, le responsable du projet nous déclare : « *Je leur ai juste fait une remarque en leur disant qu'ils étaient les seuls sur la place de Lyon à ne pas avoir signé. Et qu'on le ferait savoir !* ».

d'information¹, reconfigure les relations entre acteurs publics et privés, comme le souligne le responsable de ce projet.

J'ajoute que les nouveaux modèles économiques sont perturbateurs. Perturbateurs pour les acteurs publics, dans les services, les gens vont dire : « *on abandonne une appli mobile, on va laisser ça au secteur privé parce que ce n'est pas vraiment notre rôle de gérer les applis mobiles, qui plus est dans le carcan du code des marchés publics, on ne sait pas très bien faire ça de manière réactive* ». Donc, c'est un vrai crève-cœur pour ceux qui ont porté ça. Et puis, pour le côté secteur privé, c'est tellement mieux d'avoir un seul client [la collectivité] que d'en avoir 400 000. La prise de risque n'est pas la même. Donc, c'est aussi perturbateur côté secteur privé².

Alors qu'auparavant les acteurs publics fournissaient un service d'information aux voyageurs, la mise à disposition des données recompose leur rôle puisqu'ils décident de se retirer pour laisser la main au secteur privé³. La Centrale de mobilité est ainsi envisagée comme un vaste marché permettant à des entreprises privées de développer et de vendre des services à destination d'usagers devenus leurs clients. Dénoncée par ailleurs⁴, cette marchandisation des services publics associée à l'open data s'accompagne ici de la mise en place d'un système de régulation du marché. En effet, les porteurs du projet perçoivent dans le laisser-faire du marché un risque pour la gestion de la mobilité urbaine.

On ne peut pas se permettre qu'une donnée relative à la mobilité soit saisie par un opérateur privé qui a vocation à faire entrer plus de voitures dans des zones où l'on souhaite réduire l'usage des véhicules⁵.

Si on fournit des données –par exemple de trafic en temps réel – à des opérateurs qui, grâce à l'information du Grand Lyon, constatant qu'il y a des bouchons sur une voie, envoient le trafic sur d'autres voies (sans information trafic) le long desquelles il y a des écoles et des habitations, c'est totalement incohérent avec nos politiques publiques. Nous avons donc besoin d'avoir des

¹ Les services d'information aux voyageurs ne sont plus financés par la collectivité, mais par le marché : soit directement par la vente du service, soit indirectement par l'intermédiaire de la publicité.

² Intervention de Jean Coldefy lors de la table-ronde « La mobilité intelligente à l'échelle du bassin de vie », Congrès ATEC ITS France – Les rencontres de la mobilité intelligente, Paris, 28 janvier 2015.

³ L'information des voyageurs est une obligation légale pour les exploitants de réseau de transport. Ils continuent donc de fournir des services d'information monomodaux. Le Grand Lyon a également mis en place un site internet de calcul d'itinéraire multimodal. Mais, ce service ne sera pas développé sur téléphone mobile afin de ne pas entrer en concurrence avec les acteurs privés.

⁴ J. Bates, « “This is what modern deregulation looks like”: Co-optation and contestation in the shaping of the UK's Open Government Data Initiative. », art cit.

⁵ Interview de Timothée David à Millénaire 3, novembre 2012, http://www.millenaire3.com/fileadmin/user_upload/Interviews/Timothee_David_Donnees_et_services_RELU.pdf

garanties. Il faut dans l'exemple cité que les opérateurs s'engagent à modifier leur algorithme de calcul de navigation pour s'assurer qu'ils ne nuisent pas à l'intérêt général¹.

La collectivité est garante de la bonne gestion de l'espace public et de la mise en place d'une politique de mobilité qui respecte l'intérêt général. Or, en laissant des acteurs privés mettre en place des solutions d'information mobilité à partir de données mises à disposition, elle perd la maîtrise de ces services qu'elle régissait auparavant en propre. Comment s'assurer dès lors qu'ils correspondent à l'intérêt général défini par la collectivité ? Pour répondre à ces enjeux, le Grand Lyon a associé à la marchandisation des services d'information mobilité la création d'un système de régulation de l'utilisation des données. Celui-ci est incarné notamment par la mise en place de licences régissant l'usage possible des données². Cet instrument contractuel permet d'identifier les entreprises accédant aux données et offre au Grand Lyon un droit de regard sur l'usage des données. En dernier recours, il peut « *couper le flux*³ » d'accès aux données pour des motifs d'intérêt général. Parallèlement, un dispositif de labellisation permet aux entreprises d'utiliser la marque « OnlyMoov' » leur offrant « *une notoriété et un soutien notable à la diffusion de leurs services* »⁴. Attribué en fonction de critères précis, dont le respect est vérifié, le label est un instrument de gouvernement devant inciter les entreprises à se conformer à la politique de mobilité du Grand Lyon.

La marchandisation des services d'information mobilité ne constitue pas tant un retrait de l'acteur public que son repositionnement sur une fonction de régulation. Il ne produit plus de services, mais il définit les règles auxquelles doivent se conformer les entreprises. Ces instruments d'intervention sur le marché permettent au Grand Lyon de conserver une capacité d'orientation de la politique de mobilité métropolitaine.

¹Interview de Jean Coldefy à Millénaire 3, novembre 2012, http://www.millenaire3.com/fileadmin/user_upload/interviews/Jean_Coldefy-open_data_Lyon.pdf

² <http://data.grandlyon.com/connaitre-nos-licences> Consulté le 22 avril 2015.

³ Source : Journal de terrain, réunion avec Harald (18 mars 2013)

⁴ Communauté urbaine de Lyon – Bureau du 8 septembre 2014 – Décision n° B-2014-0330

3) Acquérir une légitimité en jouant un rôle de tiers de confiance

Ces opérations de coordination des producteurs et de régulation des utilisateurs de données participent à un travail de légitimation du Grand Lyon. Son pouvoir ne repose pas sur une règle légale lui octroyant la souveraineté exclusive du gouvernement de l'espace métropolitain de la mobilité (qui dépasse les frontières institutionnelles légitimes). L'institution doit donc s'appuyer sur d'autres formes de légitimation que le droit pour asseoir sa domination.

Les processus de centralisation et de régulation des données permettent au Grand Lyon de faire accepter son positionnement comme acteur intermédiaire assurant un rôle de tiers de confiance entre les producteurs et les utilisateurs de données. D'une part, sa capacité à réguler l'usage des données offre une garantie aux producteurs que les réutilisations n'iront pas à l'encontre de leurs intérêts d'exploitants de réseau. Ils sont alors davantage enclins à confier la diffusion de leurs données au Grand Lyon. D'autre part, s'il ne centralisait pas et ne rendait pas commensurable l'ensemble des données pour faire exister l'espace de la mobilité métropolitaine, les entreprises n'auraient aucun intérêt à s'adresser au Grand Lyon. Pour assurer sa légitimité auprès de ces acteurs, il doit en outre assurer la fiabilité et la pérennité des informations qu'il leur fournit afin que les entreprises puissent développer des services viables économiquement. La première caractéristique de ce rôle de tiers de confiance est ainsi de se porter garant dans l'interaction entre producteurs et utilisateurs de données entre lesquels la confiance réciproque ne va pas de soi¹. Le second critère est de se positionner comme un tiers : un acteur neutre ne défendant ni les intérêts des producteurs, ni ceux des reutilisateurs.

Nous, Grand Lyon, on s'est senti légitime pour le faire, en considérant qu'on était le seul des partenaires du tour de table qui n'était pas à la fois juge et partie, au sens où on n'était pas exploitant d'un réseau. On n'a pas de contrat de DSP avec un exploitant particulier du réseau. Un SYTRAL, il a des enjeux de faire tourner son réseau de transport en commun, donc, il aura plutôt tendance à privilégier son réseau de transport en commun. La région, plutôt de privilégier son réseau TER, etc. Nous, on n'a pas de contrat de délégation. Notre objectif, c'est que l'agglomération fonctionne, que l'usager trouve des solutions de mobilité les plus durables

¹ Pour reprendre les termes de Luhmann, la confiance est ainsi un « mécanisme de réduction de la complexité sociale » dans une situation d'incertitude. Niklas Luhmann, *La confiance, un mécanisme de réduction de la complexité sociale*, Paris, Economica, 2006, 123 p.

possibles. Donc on s'est senti légitime pour piloter ça. [...] Et par ailleurs, pour s'assurer qu'il y ait un certain respect des politiques publiques de mobilité en particulier, autant que ce soit un opérateur public qui le fasse. Donc nous¹.

La neutralité du Grand Lyon mise en avant dans ce récit est une justification du positionnement de l'institution comme tiers de confiance entre producteurs et utilisateurs de données. En mettant en avant les intérêts de l'agglomération et des usagers plutôt que les compétences de la Métropole², ce responsable de la direction de la voirie effectue un travail de montée en généralité qui dissocie le Grand Lyon de tout intérêt particulariste³.

Garante de l'interaction et se présentant comme neutre, la Métropole de Lyon se définit comme un tiers de confiance assurant la relation entre producteurs et utilisateurs de données. Elle joue un rôle d'institution en stabilisant les pratiques, cadrant les conduites et garantissant les places de chacun. Ce rôle de tiers de confiance participe à la légitimation de sa politique de données qui dépasse ses frontières institutionnelles, tout en lui permettant de gouverner la mobilité métropolitaine (cf. Chapitre 4).

*

* *

L'analyse du cas de la Centrale de mobilité permet de pointer quelques aspects des reconfigurations de pouvoir liées aux politiques de données urbaines. 1) La gouvernance de la donnée recompose le gouvernement urbain au profit d'un acteur intermédiaire qui se positionne entre les producteurs et les utilisateurs de données. 2) La légitimité de gouvernement de cet acteur réside dans sa capacité à associer durablement les producteurs, les réutilisateurs et les données urbaines en jouant un rôle de tiers de confiance. 3) Le périmètre de gouvernement de cet acteur en situation d'intermédiaire dépasse ses frontières institutionnelles. 4) Enfin, en gouvernant les données des différents opérateurs de transport, le Grand Lyon se construit une capacité de gouvernement de la mobilité urbaine.

¹ Entretien avec Guillaume, Responsable mobilité urbaine – Grand Lyon (22 janvier 2014)

² Alors que la Métropole est autant « partie » de la politique de mobilité : elle est responsable entre autres du réseau routier, de la délégation des nouvelles infrastructures de transport (voitures et vélos en libre service) et des parkings en ouvrage. Les élus métropolitains sont également en large majorité au sein du comité syndical du SYTRAL.

³ A l'instar de tout énoncé de justification comme l'ont analysé Luc Boltanski et Laurent Thévenot. Luc Boltanski et Laurent Thévenot, *De la justification : les économies de la grandeur*, Paris, Gallimard, 1991.

Ce cas de la mobilité urbaine interroge toutefois l'autonomie de la politique de la donnée, dans la mesure où les acteurs de ce secteur d'action publique ont joué un rôle central dans les orientations de la première (cf. chapitre 2). Les revendications de la DINSI, analysées dans ce chapitre, à autonomiser la politique de la donnée portent-elles leurs fruits ? La gouvernance des données ne reste-t-elle pas qu'un moyen au service des politiques sectorielles ? La politique processuelle de la donnée est-elle réduite à être subordonnée aux politiques substantielles ? Il est encore tôt pour répondre à ces questions. Toutefois, les situations de mise à l'épreuve des associations construites par la DINSI représentent autant de contestations à sa légitimité à mener une politique de la donnée. Elles pointent toutes les difficultés, décrites tout au long de cette thèse, à détacher les données des secteurs d'action publique (des « métiers ») pour constituer un secteur à part entière.

B. Des associations mises à l'épreuve, une légitimité contestée

En l'absence de règle légale attribuant au Grand Lyon la souveraineté exclusive pour gouverner les données métropolitaines, le rapport de consentement fondé sur l'adhésion plus ou moins tacite au pouvoir exercé par le Grand Lyon reste fragile et précaire. La solidité institutionnelle résulte de la force des associations qu'il est parvenu à construire avec l'ensemble des acteurs humains et non-humains. Si ces attachements se délitent, l'édifice institutionnel, patiemment construit, de gouvernement de la donnée métropolitaine se décompose. De fait, la position du Grand Lyon est régulièrement soumise à des procès en légitimité. Le processus de légitimation doit toujours être réaffirmé et ne tient que tant que la situation répond aux intérêts des partenaires, que les producteurs et les utilisateurs de données consentent au Grand Lyon son autorité à gouverner les données. La légitimité du Grand Lyon à gouverner les données métropolitaines est ainsi régulièrement mise à l'épreuve. Les recompositions de pouvoir au profit d'une autorité intermédiaire liées à la gouvernance de la donnée métropolitaine sont ainsi loin d'être assurées.

1) Des médiations fragiles

Pour assurer son rôle d'intermédiaire entre producteurs et utilisateurs de données, la Métropole de Lyon est dépendante des médiations qu'elle a mises en place pour assurer la relation entre ces acteurs. Si ces médiations sont fragiles, le réseau qu'elle a patiemment

construit ne tient plus. La légitimité du Grand Lyon repose notamment sur son infrastructure de diffusion des données. Elle doit être suffisamment robuste pour assurer la fourniture continue des données afin de maintenir le lien entre producteurs et utilisateurs de données. *« En matière d'open data, on est très dépendant de ce socle technique. On tire notre légitimité d'une infrastructure technique. Le socle technique amène la légitimité¹. »* indique l'administratrice générale des données. Or, lors d'une manifestation d'agriculteurs à l'été 2015, le nombre de requêtes sur la plateforme s'est multiplié ce qui a conduit à une saturation de l'infrastructure. *« C'est le gros paradoxe, au moment, où la plateforme était pleinement exploitée, elle tombe. Depuis, je m'interroge sur notre légitimité à mener une politique de la donnée². »* Pour maintenir sa position d'intermédiaire de la chaîne de la donnée, la Métropole de Lyon est dans l'obligation de s'assurer que son infrastructure technique réponde aux besoins des utilisateurs. Le cas échéant sa légitimité à mener une gouvernance de la donnée est mise à l'épreuve.

2) Des producteurs non associés

La capacité du Grand Lyon à gouverner la donnée est limitée par définition aux données pour lesquelles les producteurs lui ont donné mandat pour en assurer la gestion et la diffusion. Or, les producteurs intéressés et enrôlés à la gouvernance de la donnée métropolitaine défendue par la Métropole de Lyon sont aujourd'hui restreints à quelques secteurs d'action publique. Le Grand Lyon n'est pas parvenu à traduire les intérêts de certains producteurs particulièrement récalcitrants et à les associer à sa cause comme par exemple certaines communes : *« Le Grand Lyon est un grand organisme qui pense avoir raison sur tout. [...] C'est toujours du top-down le Grand Lyon. Depuis vingt ans, ça nous tombe du ciel, Dieu nous envoie... »³*. Cet extrait d'entretien souligne le décalage entre la problématisation de l'ouverture des données par la Métropole et les intérêts de cet administrateur de données de la ville de Lyon. Il manifeste toute la difficulté du travail d'intéressement et d'enrôlement des producteurs de données.

¹ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

² Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

³ Source : Entretien avec Philippe, administrateur de données – Ville de Lyon (25 juillet 2014).

La capacité du Grand Lyon à gouverner les données varie également selon les secteurs d'activité. Les données relatives à la mobilité urbaine et les données géographiques sont les deux secteurs principaux aujourd'hui couverts par le spectre de la gouvernance des données mise en place par la direction des services informatiques. Si la DINSI souhaite réaliser un travail similaire de centralisation et de diffusion des données relatives aux politiques énergétiques et sociales, aujourd'hui, une large majorité des directions métiers de la collectivité continuent à gérer leurs données de manière autonome.

Deux remarques à ce point peuvent être avancées. En premier lieu, si les capacités de régulation de la donnée métropolitaine par la DINSI restent, pour le moment, limitées, la gouvernance de la donnée est un projet récent, dont il est difficile aujourd'hui de percevoir et d'étudier tous les effets. En second lieu, l'enjeu de gouvernance de la donnée varie selon les secteurs de politique publique. La maturité de cet enjeu diffère : s'il est très prégnant dans le secteur de la mobilité urbaine, il est par exemple moins mis en avant dans les politiques de propreté. Surtout, même si ce point mériterait d'être étudié davantage, cet enjeu semble être construit de manière différente selon les politiques publiques. Selon les secteurs, différents objectifs sont assignés à la donnée : favoriser le développement économique, encourager la transparence, renforcer l'évaluation des contrats de délégations de service public, développer de nouveaux services à destination des citoyens, assurer à la collectivité une capacité d'expertise, etc. Pour chacune de ces problématiques, les acteurs associés et les modes d'attachement ne seront pas les mêmes. Il semble donc illusoire de construire une politique métropolitaine de la donnée autonome des secteurs d'action publique. Selon les enjeux et les politiques publiques, le gouvernement de la donnée risque d'être spécifique : il n'y a pas de propriétés permanentes qui s'appliquent à toutes les données. Dès lors, plutôt qu'une politique publique autonome de la donnée, émergerait une action publique au sein de laquelle la DINSI occupe un rôle variable selon les données ou les secteurs gouvernés.

3) Des utilisateurs en dissidence

La légitimité du Grand Lyon à gouverner les données est également mise à l'épreuve par les utilisateurs qui ne souhaitent pas s'associer à sa politique d'*open data*. Or, si personne n'utilise les données, l'assemblage ne tient pas. Deux cas de figure sont envisagés : soit les

modalités d'association trahissent les intérêts des utilisateurs, soit les utilisateurs se basent sur d'autres données non gérées par la Métropole de Lyon.

Le premier cas se manifeste par exemple avec des entreprises refusant de se soumettre aux contraintes des licences d'utilisation des données. L'article 15 des licences de diffusion suscite en particulier de nombreuses controverses puisqu'il offre à la Métropole de Lyon un droit d'audit sur le système d'information du réutilisateur.

Un des problèmes au Grand Lyon, qui est très original, c'est leur licence. C'est vraiment bizarre, parce que j'étais à fond partisan, je trouvais qu'il y avait une très bonne idée derrière, mais dans la pratique, ça ne peut pas marcher si on lit vraiment ce qu'il y a écrit. IBM, c'est une entreprise de 400 000 personnes, comment on peut estimer avoir un droit de regard ? De toute façon, le Grand Lyon ne l'aura pas, ça ne marchera jamais. Il faudrait un droit d'audit limité¹.

Comme l'exprime ce chef de projet de l'entreprise IBM, la traduction des intérêts des entreprises réutilisatrices est devenue trahison. Alors qu'il était partisan de la problématisation initiale de la Métropole de Lyon, il entre en dissidence en considérant comme abusives les contraintes que lui impose le Grand Lyon. Plusieurs critiques similaires ont été formulées à l'égard de ces licences², par des utilisateurs potentiels des données, qui ne participeront pas à la politique d'*open data* du fait des règles imposées. Conçues comme un instrument d'association des réutilisateurs, les licences deviennent un facteur de dissociation.

Le second cas de figure révèle la concurrence subie par la Métropole de Lyon face à des producteurs alternatifs de données auxquels les entreprises préfèrent s'associer. D'autres acteurs se positionnent en effet dans une situation similaire à celle du Grand Lyon d'intermédiaire entre producteurs et utilisateurs de données. L'exemple le plus emblématique

¹ Source : Entretien avec Jacques, chef de projet, IBM (27 janvier 2015)

² Voici par exemple deux extraits d'email reçu par l'équipe en charge de la plateforme de diffusion de données.

« L'article 15 semble particulièrement intrusif vis-à-vis du système d'information du licencié. Les explications fournies me paraissent disproportionnées vis-à-vis des informations mises à disposition : permettre d'accéder à un système d'information à un tiers est une demande extrêmement disproportionnée alors que nous parlons d'horaires théoriques de transport en commun. Fournir des informations suite à questions dans le cadre d'un audit me semble beaucoup plus d'adapté et surtout moins intrusif que fournir l'accès à un système d'information. » (2 mars 2015)

« Les conditions de votre licence « Engagée » sont inacceptables. [...] La non définition du terme audit vous donne trop de pouvoir. L'ensemble de la licence « Engagée » est inadapté à l'utilisation à une utilisation ouverte et libre, inadapté à ce que vous appelez « innovation ». Dommage, ça aurait pu être bien, mais ça ne l'est pas. Et vous vous privez de nombreuses participations. » (24 février 2015)

est celui des « plateformes », ces entreprises de l'économie numérique qui concentrent des données et les diffusent selon différentes modalités. Indépendantes de la Métropole de Lyon, celle-ci se retrouve dépourvue de toute « prise » pour agir sur ces entreprises. Dans le secteur de la mobilité, l'entreprise Google effectue par exemple un travail similaire de centralisation et d'association de données pour proposer des services d'information à ses utilisateurs¹. Affranchie des fournisseurs publics de données (cf. Chapitre 4), elle dispose de données recueillies directement auprès des utilisateurs des réseaux de transport². Ces traces numériques lui permettent d'agir sur l'espace urbain de manière totalement autonome des acteurs publics.

Ces deux cas de dissidence des utilisateurs soulignent la fragilité de leur attachement à la chaîne de la donnée proposée par le Grand Lyon. Or, si ces liens s'effritent, la légitimité de la Métropole de Lyon à assurer la gouvernance des données métropolitaines s'amenuise.

4) Le « gouvernement à distance » des données métropolitaines par l'Union européenne et l'Etat

La légitimité du Grand Lyon à gouverner l'espace métropolitain de la mobilité est également mise à l'épreuve par la volonté du gouvernement français d'encourager la libre circulation des données. Depuis quelques mois, l'Etat multiplie les législations en ce sens. La loi « Notre » impose l'*open data* aux collectivités de plus de 3500 habitants³, la loi « Macron » aux opérateurs de transport⁴, la loi « Valter », qui transpose la directive européenne PSI Re-use, impose un principe de gratuité des informations publiques⁵, le projet de loi « Lemaire » pour une République numérique, adopté par le gouvernement le 9 décembre 2015, met en avant un principe d'ouverture des données « par défaut » des administrations et une notion de « données d'intérêt général » qui faciliterait la mise à

¹ L'analyse est centrée sur l'entreprise Google, mais elle pourrait être valable pour d'autres acteurs de l'économie numérique. Google est néanmoins le cas qui apparaît le plus fréquemment dans notre terrain d'étude.

² Ces données « crowdsourcées » sont obtenues automatiquement auprès des utilisateurs de téléphone équipés du système d'exploitation Android ou d'applications appartenant à Google (Google Maps, Waze, etc.).

³ Source : Loi n° 2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République

⁴ Source : Loi n° 2015-990 du 6 août 2015 pour la croissance, l'activité et l'égalité des chances économiques

⁵ Source : Loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public

disposition des données des délégataires de service public¹. Si ces lois peuvent être des atouts pour accompagner la politique du Grand Lyon (par exemple, en facilitant l'obtention des données des communes ou des entreprises délégataires de service public), elles limitent également l'autonomie du Grand Lyon dans la détermination de sa politique d'*open data*. Les modalités d'ouverture des données tendent à être précisées par l'Etat (formats, gratuité, etc.), remettant en cause les licences élaborées par le Grand Lyon (cf. chapitre 2). Pour tenter d' enrôler l'État à sa cause, la Métropole de Lyon met en place des opérations de lobbying par le biais de rencontres entre ses élus et les ministres du gouvernement², de la rédaction d'amendements au projet de loi et tente de mobiliser d'autres collectivités à sa cause³. Cependant, le Grand Lyon fait le constat de son impuissance actuelle à étendre son réseau et à faire du « modèle lyonnais » un point de passage obligé des politiques de la donnée.

L'histoire de la donnée, elle se déroule à mon avis au niveau européen. Il faut qu'on soit attentif à ce qui se fait au niveau européen pour être force de proposition, en tout cas être acteur. [...] Aujourd'hui, on n'est pas dans les bons cercles. Il faut qu'on s'installe dans des réseaux de façon plus visible, et que l'on porte une parole un peu différenciante, un peu novatrice⁴.

Parce qu'elle n'est pas présente dans les « bons cercles », la Métropole de Lyon ne parvient pas à acquérir une légitimité suffisante pour asseoir sa domination sur la politique de la donnée métropolitaine. Au contraire, par l'intermédiaire de ces lois, c'est l'Union européenne et l'État qui imposent leur volonté aux collectivités territoriales. Le « libéralisme informationnel⁵ », qu'ils promeuvent leur offre une capacité de « gouvernement à distance »

¹ Source : Projet de loi pour une République numérique, <http://www.republique-numerique.fr/media/default/0001/02/ce21a30ba6d31b99c71311438a172e3c547c9dca.pdf> (Consulté le 15 décembre 2015)

² Des notes sur la politique de la donnée de la Métropole de Lyon sont rédigées à destination de Gérard Collomb et de Karine Dognin-Sauze, vice-présidente à l'innovation, la Métropole intelligente et le développement numérique, afin qu'ils évoquent ce sujet lors de leurs rencontres avec respectivement le ministre de l'économie, de l'industrie et du numérique, Emmanuel Macron, et la secrétaire d'État chargée du numérique, Axelle Lemaire.

³ Notamment au travers de message dans les listes de diffusion des groupements de collectivités territoriales et au sein de l'association OpenDataFrance.

⁴ Source : Entretien avec Léa, administratrice générale des données, Grand Lyon (21 octobre 2015)

⁵ Benjamin Loveluck nomme « libéralisme informationnel » la reformulation de la théorie politique libérale fondée sur les flux d'informations au sein des réseaux informatiques.

B. Loveluck, « Internet, une société contre l'Etat ? Libéralisme informationnel et économies politiques de l'auto-organisation en régime numérique », art cit.

des collectivités¹. En imposant la diffusion des données des acteurs locaux selon des modalités particulières, ils restreignent le champ des possibles des collectivités. Parce qu'elle ne parvient pas à enrôler l'État à sa cause par ses opérations de lobbying, la Métropole de Lyon se voit ainsi contrainte à repenser sa politique de la donnée pour la faire correspondre au cadre défini par l'Etat.

*

* *

La légitimité du Grand Lyon, et de sa direction des services d'information, à intervenir sur ce sujet et à piloter la gouvernance de la donnée est loin d'être assurée. L'enjeu de la donnée n'est aujourd'hui l'exclusivité de personne. L'Union européenne, l'État, les collectivités territoriales, les différents services de ces collectivités, les entreprises publiques et privées traitent cet enjeu, selon des intérêts divers, qui coexistent, s'entremêlent et entrent parfois en conflit. En outre, la question de la donnée surgit au sein de nombreuses politiques publiques, c'est pourquoi il est difficile de construire une politique publique de la donnée métropolitaine autonome des « métiers » : la donnée n'est qu'une dimension intégrée dans les politiques sectorielles. Les directions du Grand Lyon insèrent ce volet « donnée » dans leurs politiques publiques de leur propre chef, sans que la DINSI, censée piloter la gouvernance de la donnée n'intervienne directement. Par exemple, la direction de l'eau a intégré cette question lors du renouvellement du contrat de délégation de service public. Indépendamment des services informatiques du Grand Lyon, elle a mis en avant ce nouvel aspect, exigé l'accès aux données du délégataire, défini des formats et des protocoles d'échange de données, etc. La capacité du Grand Lyon, et de la DINSI, à piloter la gouvernance de la donnée métropolitaine, dépendra de sa capacité à étendre et stabiliser son réseau, c'est-à-dire à enrôler des acteurs hétérogènes en prenant en compte les intérêts de chacun d'entre eux. Dès lors, la gouvernance de la donnée métropolitaine devrait prendre la forme d'une action publique à géométrie variable selon les données et les secteurs gouvernés.

¹ Renaud Epstein, « Gouverner à distance. Quand l'Etat se retire des territoires. », *Esprit*, 2006, n° 11, p. 96-111.

V. Conclusion

Ce chapitre a débuté par une incertitude sur ce qu'était la gouvernance de la donnée métropolitaine. A l'issue de celui-ci, une définition de la gouvernance de la donnée peut être avancée : gouverner la donnée, c'est réguler la circulation de l'information. Cela suppose de maîtriser les flux de données et de les orienter pour, au choix, faciliter ou restreindre leurs usages. L'enjeu est ainsi d'associer ou de dissocier la donnée à certains acteurs. Pour ce faire, gouverner la donnée, c'est également assurer la coordination d'acteurs aux intérêts hétérogènes et parfois contradictoires. Gouverner la donnée, c'est donc réduire les incertitudes dans la chaîne de la donnée à l'aide d'un ensemble de médiations venant la solidifier en associant de manière durable données, producteurs et utilisateurs.

L'histoire du gouvernement de la donnée métropolitaine reste largement à écrire. Ce chapitre en présente les prémices qui ont émergé à partir de la mise en place du projet *open data*. Il s'attache à étudier la gouvernance des données « telle qu'elle se fait¹ », sans appliquer a priori un modèle normatif, mais en analysant comment les acteurs investissent cet enjeu et tentent de réguler la circulation des données métropolitaines. En suivant cet enjeu au fil des incertitudes et de ses traductions successives, on assiste à la difficile élaboration d'une politique publique de sa formalisation à sa légitimation en passant par son institutionnalisation et son instrumentation.

Comme tout « problème » social, la donnée n'est pas un enjeu naturel qui s'impose à la Métropole de Lyon. La construction de la donnée métropolitaine comme un problème nécessitant une gouvernance spécifique pour le résoudre est le résultat de la mobilisation d'acteurs. Les membres de l'équipe du projet *open data*, ainsi que la direction des services informatiques de la collectivité, ont joué un rôle central pour déterminer ce que devait recouvrir la gouvernance de la donnée métropolitaine. Par cette activité de problématisation, ils ont également défini les rôles de chacun en recomposant le réseau sociotechnique des données au profit de la Métropole de Lyon. Cette dimension cognitive de ce que recouvre la

¹ Michel Callon et Bruno Latour (eds.), *La science telle qu'elle se fait. Anthologie de la sociologie des sciences de langue anglaise*, op. cit.

gouvernance de la donnée s'est traduite dans l'organisation administrative par son institutionnalisation au sein de la direction des services informatiques de la collectivité. La forme de gouvernement qui émerge de l'analyse de l'instrumentation et des modes de légitimation de cette politique de la donnée métropolitaine prend la figure de l'*intermédiation*. Par son positionnement en situation d'intermédiaire entre producteurs et utilisateurs, la Métropole de Lyon régule la circulation des données métropolitaines en jouant un rôle de tiers de confiance : un acteur neutre se portant garant de l'interaction entre producteurs et utilisateurs. Toutefois, si cette stratégie a fonctionné dans le cas de la mobilité urbaine, sa capacité politique à adopter un positionnement similaire au sein d'autres secteurs d'action publique est questionnée. En effet, une tension entre autonomisation et sectorialisation se manifeste dans l'étude du gouvernement de la donnée : celle-ci doit-elle être l'objet d'une politique spécifique ou n'est-elle qu'un enjeu à prendre en compte de manière différencié dans chaque politique sectorielle ? A l'issue de cette thèse, ces questions restent ouvertes et constituent autant de pistes de recherche à explorer.

Le modèle du gouvernement de la donnée constitue une autre piste de recherche stimulante. D'une part, l'analyse de la figure de l'intermédiation qui caractérise la gouvernance de la donnée promue par la direction des services informatiques de la Métropole de Lyon mériterait d'être approfondie. Quelles sont ses conditions de félicité ? Différents modèles d'intermédiation apparaissent-ils selon les acteurs ou les secteurs d'action publique ? En quoi l'intermédiation recompose-t-elle les politiques sectorielles ?

Cette figure de l'intermédiation peut être rapproché du modèle de la « captation¹ », théorisé par Benjamin Loveluck : un dispositif de recentralisation des flux de données par un acteur unique, qui les redistribuent selon différentes modalités. Cette stratégie de la *mise en relation* est particulièrement prégnante dans le cas de la Centrale de mobilité mis en œuvre par le Grand Lyon. Ce modèle, que la Métropole de Lyon souhaiterait appliquer dans d'autres secteurs comme celui de l'énergie, induit une recomposition du pouvoir urbain au profit de cet acteur intermédiaire. Il présente toutefois un risque pour les acteurs publics : celui de

¹ B. Loveluck, « Internet, une société contre l'Etat ? Libéralisme informationnel et économies politiques de l'auto-organisation en régime numérique », art cit.

« désintermédiation » au profit d'un acteur privé – une plateforme¹ – centralisant les données et s'appropriant ce pouvoir de mise en relation.

D'autre part, il me semble que l'intermédiation ne constitue qu'une des formes possibles du gouvernement de la donnée. En poursuivant avec Benjamin Loveluck, deux autres formes de coordination de l'action caractéristiques du « libéralisme informationnel » pourraient s'appliquer au gouvernement de la donnée métropolitaine. La première est la « dissémination ». S'inspirant de l'organisation originelle du réseau Internet², elle correspond à une logique distribuée où les asymétries de pouvoir seraient aplanies. Le gouvernement des données s'effectuent à partir d' « organisations autonomes décentralisées³ ». La coordination entre les entités est réalisée par l'intermédiaire de protocoles techniques (le *block chain*). En combinant les logiques de pairs à pairs (*peer-to-peer*) et les techniques de cryptographie, l'exemple du Bitcoin fait émerger la possibilité d'un remplacement des institutions financières par un dispositif de suivi des transactions entre individus⁴. Ce système distribué est expérimenté au Honduras pour gérer les données cadastrales⁵ et est à l'étude au Royaume-Uni pour être appliqué aux données de référence⁶.

La troisième forme de gouvernementalité du libéralisme informationnel mise en avant par Benjamin Loveluck est « l'auto-institution ». Les logiciels libres, et leur dérivée comme par exemple l'encyclopédie Wikipédia, correspondent à cette forme d'organisation, qui repose sur des dispositifs collaboratifs d'échange et des droits communautaires de propriété intellectuelle. Ce principe des *communs*⁷ appliqué à la donnée prendrait la forme d'une

¹ Tarleton Gillepsie, « The politics of “platforms” », *New Media & Society*, 2010, vol. 12, n° 3, p. 347-364.

² Pour une histoire du réseau Internet, voir Janet Abbate, *Inventing the Internet*, Cambridge Mass., MIT Press, 1999, 268 p.

³ Primavera De Filippi et Raffaele Mauro, « Ethereum: the decentralised platform that might displace today's institutions », *Internet Policy Review*, 25 août 2014.

⁴ Satoshi Nakamoto, « Bitcoin: A Peer-to-Peer Electronic Cash System », 2008.

⁵ Gertrude Chavez-Dreyfuss, « Honduras to build land title registry using bitcoin technology », *Reuters*, 15 mai 2015.

⁶ Paul Downey, *Registers: authoritative lists you can trust*, <https://gds.blog.gov.uk/2015/09/01/registers-authoritative-lists-you-can-trust>, 9 janvier 2015, (consulté le 15 décembre 2015).

⁷ Elinor Ostrom, *La gouvernance des biens communs. Pour une nouvelle approche des ressources naturelles.*, Bruxelles, De Boeck, 2010, 301 p.

gouvernance de la donnée reposant sur un système de régulation¹, instauré par une communauté (de producteurs, d'utilisateurs, etc.), visant à préserver la ressource commune en empêchant son appropriation exclusive. *Open Street Map*, présenté dans le chapitre précédent, s'inscrit dans ce modèle politique de gouvernement de la donnée. Il s'agit ainsi d'une forme d'auto-organisation d'un collectif, à la taille variable, réuni autour d'un projet coopératif.

Ces trois formes de coordination de l'action présentent des options politiques distinctes de gouvernement de la donnée métropolitaine. Elles peuvent toutefois se combiner dans des assemblages variés selon les données et les acteurs en jeu pour chacun des secteurs d'action publique. Dans chacun de ces cas, la place et le rôle de la collectivité restent largement à définir. Selon Benjamin Loveluck, la *captation*, la *dissémination* et l'*auto-institution* sont trois formes d'auto-organisation² qui visent à se passer de toute autorité politique. Toutefois, les acteurs publics peuvent s'emparer de ces modèles, les promouvoir ou les réfréner. Dans le secteur de la mobilité, la Métropole de Lyon tente de se positionner comme intermédiaire sur l'échelle de son territoire, tout en craignant les acteurs privés lorgnant sur cette situation monopolistique. A contrario, d'autres collectivités (Paris, Toulouse, Rennes, etc.) font le choix de promouvoir un modèle inspiré des *communs* en empêchant, par le biais de licences, l'appropriation exclusive de leurs données, sans disposer toutefois d'une communauté s'appropriant cette ressource commune.

¹ Pour une analyse des systèmes de régulation des communautés numériques, voir notamment Nicolas Auray, « Le sens du juste dans un noyau d'experts. Debian et le puritanisme civique » dans Serge Proulx, Françoise Massit-Folléa et Bernard Conein (eds.), *Internet, une utopie limitée. Nouvelles régulations, nouvelles solidarités*, Québec, Presses de l'université de Laval, 2005, p. 71-94 ; Dominique Cardon et Julien Levré, « La vigilance participative. Une interprétation de la gouvernance de Wikipédia », *Réseaux*, 2009, vol. 2, n° 154, p. 51-89.

² L'auto-organisation désigne la capacité d'un système à se gouverner lui-même.

B. Loveluck, *Réseaux, libertés et contrôle*, op. cit., p. 19.

Conclusion générale

Ce que la mise en circulation des données fait au gouvernement urbain

Au travers de l'analyse de la politique d'*open data* de la Métropole de Lyon, l'ambition de cette thèse était de suivre au plus près les données pour analyser ce que leur mise en circulation fait au gouvernement urbain. La description de cinq étapes de la chaîne des données a mis au jour une tension entre attachement et détachement : pour mettre en circulation les données, il faut parvenir à les détacher de leurs liens initiaux ce qui nécessite leur attachement à de nouveaux utilisateurs. Or, multiples, ces modalités de détachement et d'attachement mettent en évidence le pluralisme des politiques de données urbaines. Après avoir résumé, pour chacune de ces étapes, les modalités de composition avec les données, j'exposerai quelques points de conclusion mettant en exergue ce que la thèse apporte au débat sur l'*open data*, à l'analyse du gouvernement de la ville, avant d'esquisser des pistes de recherche future pour prolonger cette approche de sociologie politique des données urbaines.

Le premier chapitre s'est attardé sur les modalités de production des données pour mettre en évidence les attachements des données urbaines. Alors que les données sont souvent perçues comme des entités immatérielles, neutres, autonomes, aisément duplicables et circulables, il s'avère que les données sont au contraire constituées d'attachements. Elles sont attachées à des « métiers », à des systèmes d'information, à des applications, à des opérateurs, à des modèles économiques, à des modes de faire, à des cultures particulières, etc. Il est impossible de penser la donnée indépendamment du réseau sociotechnique que forme l'ensemble de ces éléments. Mais les données sont également productrices d'attachements. La comparaison des propriétés des données « métiers » et des données de référence met en avant deux modalités d'attachement particulières qui produisent des agencements spécifiques. Les premières attachent des « métiers » à une représentation fonctionnelle de l'espace urbain tandis que les secondes lient des acteurs hétérogènes autour d'une représentation commune de l'espace urbain, donnant consistance conjointement à l'institution communautaire et à son territoire.

Adoptant une perspective généalogique, le deuxième chapitre retrace la problématisation de l'*open data* au sein de la communauté urbaine de Lyon. La politique d'ouverture des données n'a pas une seule origine linéaire, elle est le résultat d'une constellation de valeurs, de visions et de pratiques plurielles qui conduisent différents acteurs à parler d'ouverture de données de manière contingente. Au travers d'une succession

d'épreuves, la politique d'ouverture des données est traduite pour aligner les acteurs pluriels autour d'une version partagée du problème public qui réponde aux intérêts du Grand Lyon. En inscrivant la politique d'ouverture des données dans une perspective de développement économique, cette phase de problématisation redéfinit également la donnée et ses utilisateurs associés. D'instrument de politique publique, les données deviennent une ressource économique dont la circulation est créatrice de valeur pour les entreprises les réutilisant.

Le troisième chapitre est consacré au processus d'ouverture des données. La mise en circulation des données nécessite le détachement de leur environnement initial. Cela passe par une série d'épreuves au cours desquelles ces associations sont questionnées au regard de l'attachement des données à de nouveaux utilisateurs. Tout processus de détachement s'accompagne en effet d'un processus conjoint d'attachement. Trois épreuves principales de diffusibilité des données ont été identifiées. L'épreuve d'identification catégorise ce qu'est une « donnée candidate ». La publicisation fait émerger la « donnée publiable ». Enfin, l'extraction précise ce qu'est une « donnée ouverte ». La donnée est modifiée tout au long de la chaîne de diffusion pour finir par se stabiliser comme une entité diffusable. Ce processus est réversible : des données non catégorisées comme telles peuvent le devenir, et, inversement, des données ouvertes peuvent être redéfinies comme non-diffusables.

Le chapitre suivant s'intéresse aux modalités de réutilisation des données ouvertes. En décrivant les modalités d'attachement des données à de nouveaux utilisateurs, il met en avant le pluralisme des politiques de réutilisation. Elles sont regroupées dans trois catégories selon les opérations qu'elles font subir aux données : la consolidation, l'homogénéisation et l'articulation. La consolidation consiste à ajouter un ensemble d'opérations à la chaîne de la donnée afin d'assurer sa solidité au sein d'un nouvel environnement. L'homogénéisation est l'usage de données de sources hétérogènes pour produire un nouvel agrégat uniforme. L'articulation est la mise en relation de données diverses au travers d'une donnée-pivot. En recomposant les attachements des données, ces trois politiques produisent de nouveaux agencements qui associent des utilisateurs, une représentation spécifique de l'espace urbain et un mode d'action particulier sur celui-ci, influant en cela sur le gouvernement de la ville.

Le dernier chapitre analyse la construction d'une politique métropolitaine de la donnée qui vise à réguler les flux de données. Par sa mise en circulation, la donnée est en effet devenue un enjeu de gouvernement urbain. Au travers d'opérations de formalisation, d'institutionnalisation, d'instrumentation et de légitimation, la direction des services informatiques du Grand Lyon tente d'ériger la donnée en politique publique autonome dont elle assurerait le pilotage à l'échelle de la Métropole. Toutefois, les difficultés rencontrées par cette direction pour réguler la circulation des données révèlent une tension latente entre autonomisation et sectorialisation de cet enjeu de la donnée.

I. Des politiques plurielles d'open data

Le premier apport de la thèse est d'enrichir les réflexions sur l'*open data*. Récentes, les politiques d'ouverture de données ont été l'objet de peu de travaux analysant, au delà du discours des acteurs, les pratiques de diffusion de données. L'insertion au sein de la communauté urbaine de Lyon m'a permis de suivre la politique d'ouverture des données *en train de se faire*. Ce point de vue ethnographique permet de sortir du débat normatif auquel est trop souvent cantonné l'*open data*, entre ses partisans qui pointent le renouveau démocratique et ses détracteurs qui dénoncent une marchandisation des services publics. Cette thèse enrichit ce débat en explicitant le pluralisme des médiations et des formes d'attachement qui font émerger, de l'ouverture des données, une multiplicité d'agencements.

D'une part, contrairement au postulat du libéralisme informationnel selon lequel les données existent et peuvent aisément circuler, la thèse met en évidence les multiples attachements des données qui ne préexistent pas en tant qu'entités autonomes aux processus d'ouverture. Toute donnée est indissociable du réseau sociotechnique qui la constitue ; pour la mettre en circulation, il est nécessaire de recomposer ce réseau en déliant certaines associations et en constituant de nouvelles. Loin d'être joué à l'avance, le processus d'ouverture des données est le résultat d'une série d'épreuves, au résultat toujours incertain, au cours desquelles les caractéristiques des données, des producteurs, des utilisateurs, sont jugées et redéfinies par les acteurs. Ces différentes *épreuves de diffusibilité* conduisent ainsi à recomposer le réseau des données afin qu'elles soient considérées comme « diffusables ». Cette modification du réseau sociotechnique transforme la donnée. A mesure que la donnée

circule d'une épreuve à l'autre, elle diffère par le réseau qu'elle déploie. Ainsi, la donnée comme entité stable et immuable est une fiction : il n'y a donc pas une, mais des données dont les caractéristiques évoluent parallèlement aux attachements qui les constituent.

D'autre part, alors que les partisans de l'*open data* prônent une mise à disposition des données sous des conditions strictes (libre accès, gratuité, format ouvert, etc.), la thèse met en avant la pluralité des modalités de mise en circulation des données. Selon les données, les secteurs d'action publique, les objectifs poursuivis, les infrastructures disponibles, ces principes sont traduits localement au travers des médiations qui assurent la mise à disposition des données. C'est toute l'essence de la politique que de composer un monde commun entre les intérêts en présence, et il serait réducteur de vouloir imposer une seule et unique politique de libre circulation des données. Au contraire, l'art de gouverner les données réside dans la capacité à composer entre différentes approches de régulation des flux de données. Plutôt que de les considérer de manière uniforme, les politiques de mise en circulation des données doivent ainsi être envisagées au pluriel. La mise en évidence de ce pluralisme contribue à raviver le débat démocratique en explicitant l'étendue des choix possibles et évite ainsi l'imposition d'une « subpolitique¹ » ancrée dans une infrastructure technique uniforme.

II. Ce que la mise en circulation des données fait au gouvernement urbain

Au delà de l'enrichissement du débat sur l'*open data*, cette thèse contribue aux réflexions sur le gouvernement des villes. Je souhaite revenir dans cette conclusion sur une question qui traverse cette thèse : que fait la mise en circulation des données au gouvernement urbain ? En effet, si l'*open data* constitue le parangon de la mise en circulation des données urbaines, ce dernier processus est croissant au sein des villes à l'ère du numérique. Sous le vocable de la *smart city*, les flux d'information sont pensés comme la solution à l'ensemble des problèmes sociaux. Ces discours aux inspirations cybernétiques traduisent la mise en œuvre du libéralisme informationnel à l'échelle des villes. De nouvelles modalités de

¹ Ulrich Beck, *La société du risque : Sur la voie d'une autre modernité*, Paris, Flammarion, 2008, 521 p.

régulation de l'espace urbain apparaissent, ce qui contribue à constituer la donnée comme un enjeu de gouvernement et questionne les frontières institutionnelles et territoriales.

A. La mise en circulation des données fait de la donnée un enjeu de gouvernement

Durant les trois ans de mon étude (2012-2015) au sein de la communauté urbaine de Lyon, j'ai constaté que le sujet majeur de préoccupation de mes interlocuteurs dérivait de l'*open data* vers la gouvernance des données. Je me suis longtemps interrogé sur les raisons de cette évolution pour arriver à la réponse suivante : les données sont devenues un enjeu de gouvernement à partir du moment où elles ont été mises en circulation. Attachées à des producteurs, conservées dans des silos « métiers », les données sont des instruments au service d'une politique publique sectorielle. Elles visent à décrire l'espace urbain pour agir dessus ou produire une connaissance pour aider à la décision. Elles sont un outil mais ne sont pas l'objet d'enjeux spécifiques.

Alors que, comme l'indique sa vice-présidente, la stratégie de ville intelligente du Grand Lyon repose sur le passage progressif « *de l'idée de "faire la ville" à celle de "faire faire la ville" »*¹, la mise en circulation des données renouvelle les modes de faire de la collectivité en s'appuyant sur des acteurs tiers qui produisent des services à partir de ses données. Toutefois, la Métropole souhaite conserver une capacité de contrôle de l'usage qui est fait de ses données. Elle doit pour cela mettre en place une stratégie de régulation des flux de données. Par le choix des modalités de mise en circulation des données (licences, formats, etc.), elle encourage l'attachement de ses données à certains utilisateurs ou au contraire en limite d'autres. Ces médiations permettent au Grand Lyon de privilégier l'émergence de certains agencements avec ses données ouvertes au détriment d'autres. Le gouvernement de la donnée contribue ainsi à renouveler les modalités de gouvernement de la Métropole de Lyon.

Au travers du gouvernement des données, c'est la capacité de l'acteur public à gouverner la ville à l'ère du numérique qui est en jeu. Alors que de plus en plus d'actions et de comportements sont régulés par des flux d'informations, si l'acteur public ne contrôle pas

¹ Source : <http://www.latribune.fr/regions/smart-cities/20141011trib2a8fd90bb/les-nouvelles-promesses-de-la-ville-intelligente.html> (11 octobre 2014) Consulté le 11 août 2015.

ces flux, il perd la maîtrise de ses politiques publiques. L'exemple du GPS cité dans le chapitre 4 est révélateur de ce risque. Régulièrement, la presse locale fait sa Une sur le rôle des GPS dans la congestion urbaine et le fameux « bouchon de Fourvière ». En effet, les applications basées sur les technologies GPS n'utilisent pas la même classification de voirie que celle du Grand Lyon et les calculs d'itinéraires déterminés par des algorithmes conduisent les automobilistes, qui suivent les indications de leur appareil, à emprunter le tunnel de Fourvière plutôt que d'autres itinéraires de substitution, au grand dam des régulateurs de la Métropole de Lyon. Plusieurs réunions avec les fabricants d'applications de GPS ont eu lieu pour tenter de faire évoluer leurs données, sans succès pour le Grand Lyon. Seule une action sur l'infrastructure routière, par le déclassement de l'autoroute en boulevard urbain, règlera cette situation. Dans l'incapacité de réguler les flux de données, la collectivité perd la maîtrise de sa politique de régulation du trafic routier.

Cet exemple met en avant tout l'enjeu que constitue la régulation des flux de données pour conserver la maîtrise des politiques publiques. Or, de quels moyens le Grand Lyon dispose-t-il et quelle est sa légitimité à réguler des flux de données sur lesquelles il n'a pas la main ? Indépendamment des pouvoirs publics, des acteurs peuvent réguler la ville par le biais de flux informationnels. Le gouvernement des données métropolitaines est une piste de recherche à explorer qui mériterait des travaux comparant des villes, des institutions, des secteurs d'action publique et des natures de données pour mettre en évidence les capacités de régulation des acteurs publics.

B. La mise en circulation des données questionne les frontières sectorielles, institutionnelles et territoriales

Le premier chapitre de cette thèse a mis en évidence l'attachement des données à un métier ou une institution produisant un certain type d'agencement du gouvernement de la ville. La mise en circulation des données vise précisément à défaire ces attachements pour rompre les silos sectoriels de données et produire une *digital skin*¹. Les promesses de la ville intelligente résident dans cette transversalité qui fait fi de ces frontières. Détachées, les

¹ C. Rabari et M. Storper, « The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data », art cit.

données ne sont plus en congruence avec les découpages administratifs et les territoires tels qu'ils sont institutionnalisés.

De fait, l'analyse des réutilisations de données illustre l'émergence de nouveaux agencements qui questionnent les frontières institutionnelles ou territoriales. Tout d'abord, la consolidation des données rend possible leur usage au sein d'autres environnements sociaux sans que la représentation de l'espace urbain ne soit affectée. Mais surtout, l'homogénéisation ou l'articulation produisent des agencements particuliers associant une représentation alternative de l'espace urbain à de nouvelles modalités d'action sur la ville.

Toutefois, les réutilisations supposent un travail délicat de mise en relation et de coordination de mondes sociaux hétérogènes. Cette complexité de l'attachement des données ouvertes à de nouveaux utilisateurs limite l'émergence de ces agencements qui mettent en question les frontières institutionnelles. Surtout, la mise en circulation ne fait pas disparaître les silos de données. Les données restent attachées aux métiers : elles continuent d'être produites et utilisées pour répondre aux besoins des « métiers » de la collectivité. Les silos de données ne disparaissent pas : ces nouveaux agencements s'ajoutent aux précédents.

Cette tension entre attachement et détachement des données met en évidence toute la difficulté à produire des agencements alternatifs qui dépassent réellement les secteurs ou les territoires d'action publique. Les politiques de données semblent encore majoritairement sectorielles, ce que souligne toute la difficulté de la direction des services informatiques de la Métropole de Lyon à mettre en place une politique des données autonomes des « métiers » de la collectivité (cf. chapitre 5). Plutôt que de remettre en cause les « silos » métiers, la ville numérique fait émerger des politiques de données qui varient selon les secteurs d'action publique. Pour approfondir ce résultat, il serait intéressant de comparer le secteur de la mobilité à celui de l'énergie.

Pour répondre à des objectifs de transition énergétique, le secteur de l'énergie est l'objet de profondes recompositions de gouvernance avec un mouvement conjoint de libéralisation et de décentralisation. Les métropoles, qui se sont vues attribuer de nouvelles

compétences en matière de gestion et de pilotage des réseaux énergétiques¹, œuvrent à l'élaboration de politiques énergétiques territorialisées. Les données ont une place majeure dans ces politiques. En premier lieu, les collectivités doivent diminuer l'asymétrie d'informations avec les acteurs du secteur pour parvenir à exercer ces compétences, élaborer des diagnostics territoriaux et mettre en œuvre et piloter ces politiques énergétiques territorialisées. En second lieu, le déploiement de capteurs² et de compteurs communicants³ offre de nouvelles sources de données qui permettent d'optimiser la gestion des réseaux et d'agir plus finement sur la consommation énergétique. Ces deux points soulèvent des enjeux de collecte, d'échange, de mutualisation, de consolidation et de mise à disposition de données hétérogènes similaires à ceux décrits dans le secteur de la mobilité urbaine. Verra-t-on des politiques de données identiques émerger ? Quel rôle sera dévolu aux collectivités ? Dans quelle mesure les données conduiront-elles à la recomposition de la gouvernance énergétique locale ? Participeront-elles à une « déssectorialisation » de l'énergie et à l'intégration de l'enjeu de la transition énergétique au sein des politiques de mobilité et d'aménagement de l'espace ?

III. Les pistes de prolongement de la recherche : pour une sociologie politique des données urbaines

Au travers de cette thèse, j'ai suivi une approche de sociologie politique de la donnée urbaine qui croise les analyses internalistes et externalistes pour étudier les recompositions des modalités de gouvernement des espaces urbains. Suivre la chaîne de la donnée, c'est mettre en avant différents agencements entre des acteurs, des modes de gouvernement et des représentations de la ville. Cela offre un nouveau regard sur les modalités de coordination entre les acteurs, les secteurs, les institutions et les territoires du gouvernement urbain. Cette

¹ Notamment au travers des lois n°2014-58 du 27 janvier 2014 de modernisation de l'action publique territoriale et d'affirmation des métropoles (Maptam), n°2015-991 du 7 août 2015 portant nouvelle organisation territoriale de la République (Notre) et n°2015-992 du 17 août 2015 relative à la transition énergétique pour la croissance verte.

² Pour améliorer l'efficacité des réseaux, sont mis en œuvre des *smart grids*, des réseaux énergétiques « intelligents » dotés de capteurs devant permettre de faciliter leur pilotage.

³ Enedis (anciennement ERDF) déploie actuellement le compteur Linky qui sera installé dans 35 millions de foyers d'ici 2021 ; GRDF va équiper 11 millions de foyers de son compteur communicant Gazpar de 2016 à 2022.

approche mériterait d'être appliquée aux nouvelles formes de quantification indépendantes des acteurs publics. De plus, la sociologie politique de la donnée me paraît être une entrée particulièrement stimulante pour questionner la notion de territoire par la mise en évidence d'agencements spatiaux pluriels.

A. Nouvelles formes de quantification et nouveaux acteurs de gouvernement urbain

Cette thèse est centrée sur les données publiques. Pourtant, à l'ère du numérique, de nombreuses données échappent aux acteurs publics. La baisse des coûts des capteurs, du stockage et du calcul a conduit à une explosion de la production de données. En permettant à des acteurs divers d'agir sur l'espace urbain, ces nouvelles modalités de quantification interrogent le gouvernement urbain et la place des acteurs publics dans celui-ci. Deux pistes de réflexion mériteraient d'être poursuivies : les données personnelles et les traces captées par les plateformes.

Au travers de leurs usages de services numériques, les individus produisent des données qui ne cessent de croître en volume et en diversité. Ces données, qualifiées de données personnelles, sont aujourd'hui majoritairement conservées au sein de silos institutionnels, privés ou publics. Toutefois, les évolutions réglementaires à l'échelle européenne et nationale visent à instaurer des principes d'autodétermination des données, au travers par exemple, d'un droit à la portabilité des données, devant faciliter aux individus la migration d'un service à un autre. Ces tendances pointent l'application de logiques de mise en circulation des données personnelles desquelles peuvent émerger différentes politiques de données. L'utilisateur pourrait devenir le courtier de ses propres données, libre de ses choix pour les valoriser. Si cette solution s'inscrit dans une perspective d'*empowerment*, elle laisse également transparaître des risques de renforcement des inégalités entre les individus, dont la valorisation des données sera inégale. Surtout, en considérant que les données personnelles sont des biens individuels sur lesquels chacun dispose d'un droit de propriété, elle souligne une évolution majeure de la conception de la vie privée qui n'est plus pensée comme un bien collectif régie par des normes communes visant à la protéger¹ mais comme une liberté individuelle dont les contours

¹ La protection des données personnelles est alors envisagée comme un droit fondamental.

sont définis par chaque individu¹. Une autre voie serait de considérer les données personnelles comme des communs, gérées collectivement par l'intermédiaire d'une gouvernance partagée de cette ressource qui deviendrait inappropriable. Enfin, une autre piste résiderait dans la régulation collective des données personnelles, par des acteurs publics ou privés assurant un rôle de tiers de confiance entre les individus et les entreprises utilisatrices de leurs données². Esquissées très sommairement, ces trois pistes de politiques des données personnelles mises en circulation laissent entrevoir la pluralité des modalités de composition avec les données personnelles, au sein desquelles les agencements entre acteurs, les modes de régulation et les conceptions politiques diffèrent. A l'échelle urbaine, si les collectivités n'ont jamais été au cœur de la protection de la vie privée et de la régulation des données personnelles, elles pourraient demain jouer un rôle plus important en encourageant telles ou telles politiques des données personnelles.

Un second champ de recherche à explorer est celui formé par les acteurs de l'économie numérique, constitués en plateformes situées en position d'intermédiaire entre producteurs et utilisateurs de données. En parvenant à capter des traces numériques laissées par leurs utilisateurs, ces entreprises sont capables de proposer des services et d'agir sur l'espace urbain de manière totalement autonome des acteurs publics. Ces évolutions en matière de production de données sont susceptibles de contribuer à recomposer les gouvernements urbains et à questionner la souveraineté des collectivités locales. Comment agir sur des acteurs qui font fi de toutes frontières territoriales et institutionnelles ? Cette interrogation pointe les nouvelles formes de partenariat entre acteurs publics et privés qui peuvent émerger au travers des données urbaines. Au-delà de la capacité des acteurs publics à conserver la maîtrise sur leurs politiques publiques, se pose la question de la légitimité démocratique de ces services opérés

¹ Comme le pointe Dominique Cardon : « *La vie privée a longtemps été pensée comme un bien collectif à partir duquel était érigé un ensemble de normes communes à tous, mais aussi de valeurs partagées par l'ensemble de la société. [...] Cette définition univoque de la vie privée est aujourd'hui fragilisée par le souci des individus d'en définir eux-mêmes la teneur et de ne pas laisser à d'autres le soin de ses contours. Construite comme une protection, la vie privée est de plus en plus vécue comme une liberté. Les utilisateurs d'internet souhaitent mieux contrôler ce qu'ils acceptent de rendre public ou de confier à d'autres.* »

D. Cardon, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, op. cit., p. 78.

² C'est le sens de l'expérimentation MesInfos mise en œuvre par la FING à laquelle participe la Métropole de Lyon (<http://mesinfos.fing.org>). Cette dernière pourrait envisager de se positionner comme tiers de confiance entre les citoyens et les fournisseurs de services urbains utilisant des données personnelles.

par des plateformes commerciales. Quels seront les espaces de débat et de représentation politique au sein de ces nouveaux agencements urbains ?

B. Le pluralisme des agencements territoriaux

L'analyse des modes de production des données publiques a mis en évidence l'existence de différents agencements territoriaux. Attachées à un métier ou une institution, représentant l'espace urbain selon des conventions prédéfinies, les données véhiculent une représentation particulière de l'espace urbain selon des conventions prédéfinies. Le premier chapitre a mis en évidence le fait que, selon les attachements qui leur donnent corps, les données produisent des agencements spatiaux spécifiques. J'ai identifié deux formes d'agencement : un agencement fonctionnel qui émerge de l'attachement entre un service « métier » et une représentation de l'espace urbain à visée opérationnelle ; et un agencement territorial qui fait tenir ensemble des acteurs hétérogènes autour d'une représentation commune de l'espace urbain.

La mise en circulation des données interroge ces agencements spatiaux. D'une part, elle rompt l'association exclusive entre l'acteur public et l'espace sur lequel il exerce son pouvoir. La mise à disposition des données offre à des acteurs externes la possibilité d'agir sur le territoire de l'institution publique au risque d'une marginalisation éventuelle de cette dernière. D'autre part, la dynamique des flux de données conduit à une recomposition des échelles de gouvernement. La mise en relation de données hétérogènes fait émerger de nouvelles représentations qui recomposent les frontières gouvernementales que celles-ci soient institutionnelles ou territoriales. A l'instar de la Métropole de Lyon qui régule la mobilité au-delà de ses frontières institutionnelles par l'articulation de données, une mosaïque d'agencements spatiaux apparaît, complexifiant les échelles territoriales, toujours plus variées, enchevêtrées et parfois contradictoires, avec lesquelles une institution peut jouer pour gouverner.

Cette remise en question de la distinction interne/externe invite à penser différemment notre modèle de souveraineté basé sur le couplage d'une institution et d'un territoire. Comment penser le pouvoir politique indépendamment d'un espace délimité par des frontières sur lequel s'exerce une autorité exclusive ? Cette interrogation invite à explorer

d'autres formes d'agencements associant une institution à des espaces¹. Au travers du territoire, elle questionne également l'État moderne dans la mesure où la théorie politique associe étroitement le territoire, envisagé comme un espace délimité par des frontières sur lequel un pouvoir exclusif est exercé, à l'État moderne, qui exerce sa souveraineté sur une population vivant sur un territoire délimité par des frontières².

Quel est alors le rôle de l'acteur public s'il n'a plus l'autorité exclusive sur un territoire ? Comment une institution politique parvient-elle à partager le pouvoir de gouvernement ? Pour résoudre ces enjeux de coordination entre des acteurs agissant sur des espaces qui se chevauchent et qui se jouent des échelles de régulation politique, il faut inventer de nouvelles formes de diplomatie. Les données participent à la composition de ces collectifs. Au travers des politiques de données urbaines décrites dans cette thèse, la collectivité dispose d'une modalité de régulation lui permettant de jouer avec la pluralité des agencements spatiaux. Toutefois, la question démocratique se pose : comment inventer des arènes de discussion et des modalités de représentation non basées sur une logique territoriale ?

¹ Dans cette perspective, Dominique Boullier questionne cette notion de territoire saisi par le numérique en proposant trois agencements : le territoire à agencement topographique, le territoire à agencement topologique et le territoire à agencement chronologique. Dominique Boullier, « L'écume des territoires » dans Marta Severo et Alberto Romele (eds.), *Traces numériques et territoires*, Paris, Presses des Mines, 2015, p. 113-134.

² En témoigne la définition canonique de l'État par Max Weber qui reflète cette relation étroite entre État, souveraineté et territoire : « *Ce qui est formellement caractéristique de l'Etat contemporain, c'est une réglementation administrative et juridique, modifiable par des lois, d'après laquelle s'oriente l'entreprise de l'activité de groupement de la direction administrative (également règlementée par des lois) et qui revendique une validité non seulement pour les membres du groupement – qui y sont en substance incorporés par naissance – mais aussi, dans une large mesure, pour toute l'activité qui se déroule dans les limites du territoire qu'il domine (par conséquent conformément à l'institution territoriale). Il se caractérise en outre par le fait qu'il existe de nos jours de violence « légitime » que dans la mesure où l'ordre étatique le permet ou la prescrit. [...] Cet aspect du monopole de la violence réservé à la domination étatique est une caractéristique aussi essentielle de sa condition présente que son caractère d' « institution rationnelle » ou d' « entreprise » continue. » In Max Weber, *Economie et société*, Paris, Pocket, 1995, p. 99.*

Sources

IV. Bibliographie

ABBATE Janet, *Inventing the Internet*, Cambridge Mass., MIT Press, 1999, 268 p.

ABITEBOUL Serge, *Sciences des données : De la logique du premier ordre à la Toile*, Paris, Fayard (coll. « Collège de France »), 2012, 112 p.

AGENCE FRANÇAISE POUR L'INFORMATION MULTIMODALE ET LA BILLETIQUE (AFIMB), *L'open data dans le domaine du transport : analyse des premières initiatives et recommandations*, Paris, Ministère de l'Écologie, du Développement durable et de l'Énergie, 2013.

AKRICH Madeleine, « Les objets techniques et leurs utilisateurs, de la conception à l'action. » dans Bernard Conein, Nicolas Dodier et Laurent Thévenot (eds.), *Les objets dans l'action*, Raisons Pratiques., Paris, Editions de l'EHESS, 1993, p. 35-57.

AKRICH Madeleine, « Comment décrire les objets techniques ? », *Techniques et culture*, 1987, n° 9, p. 49-64.

AKRICH Madeleine, CALLON Michel et LATOUR Bruno, *Sociologie de la traduction : Textes fondateurs*, Paris, Presses de l'Ecole des Mines, 2006.

AKRICH Madeleine, CALLON Michel et LATOUR Bruno, « A quoi tient le succès des innovations ? 1 : L'art de l'intéressement », *Gérer et comprendre. Annales des Mines.*, 1988, n° 11, p. 4-17.

AKRICH Madeleine et LATOUR Bruno, « A Summary of a Convenient Vocabulary for the Semiotics of Human and Nonhuman Assemblies » dans Wiebe E. Bijker et John Law (eds.), *Haping Technology/Building Society. Studies in Sociotechnical Change*, Cambridge Mass., MIT Press, 1992, p. 259-264.

ALAM Thomas, GURRUCHAGA Marion et O'MIEL Julien, « Science de la science de l'État : la perturbation du chercheur embarqué comme impensé épistémologique. », *Sociétés contemporaines*, 2012, vol. 3, n° 87, p. 155-173.

ALLARD Laurence et BLONDEAU Olivier, *Nouvelles formes de mobilisation et d'innovation politique : le concours*, <http://www.internetactu.net/2010/06/10/nouvelles-formes-de-mobilisation-et-dinnovation-politique-le-concours/>, 6 octobre 2010, consulté le 8 avril 2015.

AMAR Georges, *Homo mobilis. Le nouvel âge de la mobilité.*, Paris, FYP, 2010.

AMIRAUX Valérie et CÉFAÏ Daniel, « Les risques du métier. Engagements problématiques en sciences sociales. », *Cultures & Conflits*, 2002, n° 47.

ANDERSON Charles W., « Political Judgment and Theory in Policy Analysis » dans Michael Levy et Edward Portis (eds.), *Handbook of Political Theory and Policy Science*, New York, Greenwood, 1988, p. 183-198.

ANDERSON Chris, « The End of Theory: The Data Deluge Makes the Scientific Method Obsolete », *Wired*, 23 juin 2008.

ANGELETTI Thomas et BERLAN Aurélien, « Les êtres collectifs en question », *Tracés*, 2015, vol. 3, n° 29, p. 7-22.

AURAY Nicolas, « Les technologies de l'information et le régime exploratoire » dans Pek Van Andel et Bourcier Danièle (eds.), *La sérendipité dans les arts, les sciences et la décision*, Éditions Hermann., Paris, 2010.

AURAY Nicolas, « Le sens du juste dans un noyau d'experts. Debian et le puritanisme civique » dans Serge Proulx, Françoise Massit-Folléa et Bernard Conein (eds.), *Internet, une utopie limitée. Nouvelles régulations, nouvelles solidarités*, Québec, Presses de l'université de Laval, 2005, p. 71-94.

AZOULAY Samuel, *De l'open data à l'open government. Le SGMAP et Etalab dans la politique de modernisation de l'action publique.*, Université de Paris II, Paris, 2014, 223 p.

BAACK Stefan, « Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism », *Big Data & Society*, 2015, vol. 2, n° 2.

BARDET Fabrice, *La contre-révolution comptable. Ces chiffres qui (nous) gouvernent*, Paris, Les Belles Lettres, 2014, 374 p.

BARTHE Yannick, *Le pouvoir d'indécision : la mise en politique des déchets nucléaires*, Paris, Economica, 2006, 239 p.

BARTHE Yannick, « Le recours au politique ou la problématisation politique « par défaut » » dans Jacques Lagroye (ed.), *La politisation*, Paris, Belin, 2003, p. 475-492.

BATES Jo, « The Strategic Importance of Information Policy for the Contemporary Neoliberal State: The Case of Open Government Data in the United Kingdom. », *Government Information Quarterly*, 2014, vol. 31, n° 3, p. 388-395.

BATES Jo, « The Domestication of Open Government Data Advocacy in the United Kingdom: A Neo-Gramscian Analysis », *Policy & Internet*, 2013, vol. 5, n° 1, p. 118-137.

BATES Jo, « “This is what modern deregulation looks like”: Co-optation and contestation in the shaping of the UK’s Open Government Data Initiative. », *The Journal of Community Informatics*, 2012, vol. 8, n° 2.

BATTY Michael, AXHAUSEN Kay, FOSCA Giannotti, POZDNOUKHOV Alexei, BAZZANI Armando, WACHOWICZ Monica, OUZOUNIS Georgios et PORTUGALI Yuval, « Smart Cities of the Future », *European Physical Journal Special Topics*, 2012, n° 214, p. 481-518.

BAUDOT Pierre-Yves, « L’informatique aux informaticiens. Sélection des acteurs, définition de l’instrument et politisation de l’informatique administrative (France, 1966-1975) » dans Gwenola Le Naour et Gilles Massardier (eds.), *L’action publique sélective*, Paris, LGDJ, 2013, p. 182-196.

BEAL Vincent, EPSTEIN Renaud et PINSON Gilles, « La circulation croisée. Modèles, labels et bonnes pratiques dans les rapports centre-périphérie », *Gouvernement et action publique*, 2015, vol. 3, n° 3, p. 103-127.

BEAUDE Boris, « Espace de la ville, espace de la carte; de l’analogie à la coexistence » dans Khaldoun Zreik (ed.), *Nouvelles cartographies, nouvelles villes HyperUrbain.2*, Paris, Europia productions, 2010, p. 15-39.

BECK Ulrich, *La société du risque : Sur la voie d’une autre modernité*, Paris, Flammarion, 2008, 521 p.

BERGERON Henri, CASTEL Patrick et DUBUISSON-QUELLIER Sophie, « Gouverner par les labels. Une comparaison des politiques de l’obésité et de la consommation durable », *Gouvernement et action publique*, 2014, vol. 3, n° 3, p. 7-31.

BERRY Gérard, *Pourquoi et comment le monde devient numérique*, Paris, Fayard (coll. « Collège de France »), 2008, 78 p.

BLONDIAUX Loïc, *La Fabrique de l’opinion. Une histoire sociale des sondages*, Paris, Seuil, 1998, 600 p.

BOLTANSKI Luc, *De la critique : Précis de sociologie de l’émancipation*, Paris, Gallimard, 2009, 294 p.

BOLTANSKI Luc et THEVENOT Laurent, *De la justification : les économies de la grandeur*, Paris, Gallimard, 1991.

BORGMAN Christine L., *Big Data, Little Data, No Data: Scholarship in the Networked World*, Cambridge Mass., MIT Press, 2015, 400 p.

BOUDREAU Kevin et HAGIU Andrei, « Platform Rules: Multi-Sided Platforms as Regulators » dans Annabelle Gawer (ed.), *Platforms, Markets and Innovation*, Northampton, Edward Elgar Publishing, 2011.

BOULLIER Dominique, *Sociologie du numérique*, Paris, Armand Colin, 2016, 352 p.

BOULLIER Dominique, « Vie et mort des sciences sociales avec le big data », *Socio*, 2015, n° 4, p. 19-37.

BOULLIER Dominique, « L'écume des territoires » dans Marta Severo et Alberto Romele (eds.), *Traces numériques et territoires*, Paris, Presses des Mines, 2015, p. 113-134.

BOULLIER Dominique, « Le client du poste téléphonique : archéologie des êtres intermédiaires » dans *Débordements. Mélanges pour Michel Callon.*, Paris, Presses de l'École des Mines, 2010, p. 41-61.

BOULLIER Dominique, *La ville événement*, Paris, Presses Universitaires de France (coll. « La ville en débat »), 2010, 112 p.

BOULLIER Dominique, « S'orienter en politique : des partis et des ennemis » dans Jacques Lolive et Olivier Soubeyran (eds.), *L'émergence des cosmopolitiques*, Paris, La Découverte, 2007, p. 100-114.

BOULLIER Dominique et CREPEL Maxime, « Biographie d'une photo numérique et pouvoir des tags. Classer/circuler », *Revue d'anthropologie des connaissances*, 2013, vol. 7, n° 4, p. 785-813.

BOURDIEU Pierre, *Le sens pratique*, Paris, Les Editions de Minuit, 1980.

BOWKER Geoffrey C., *Memory Practices in the Sciences*, Cambridge Mass., MIT Press, 2005, 280 p.

BOWKER Geoffrey C., « Biodiversity Datadiversity. », *Social Studies of Science*, 2000, vol. 30, n° 5, p. 643-683.

BOWKER Geoffrey C., BAKER Karen, MILLERAND Florence et RIBES David, « Toward information infrastructure studies: ways of knowing in a networked environment » dans Jeremy Hunsinger, Lisbeth Klastrup et Matthew Allen (eds.), *International Handbook of Internet Research*, New York, Springer, 2010, p. 97-118.

BOWKER Geoffrey C. et STAR Susan Leigh, *Sorting Things Out. Classification and Its Consequences*, Cambridge Mass., MIT Press, 1999.

BOYD danna et CRAWFORD Kate, « Critical questions for big data », *Information, Communication and Society*, 2012, vol. 15, n° 5, p. 662-679.

BROCA Sébastien, *Utopie du logiciel libre. Du bricolage informatique à la réinvention sociale*, Paris, Le Passager clandestin, 2013, 282 p.

BROUSSE Cécile, « Définir et compter les sans-abri en Europe : enjeux et controverses », *Genèses*, 2005, vol. 1, n° 58, p. 48-71.

BRUNO Isabelle, « Y a-t-il un pilote dans l'Union ? », *Politix*, 2008, n° 82, p. 95-117.

CALLON Michel, « Pour une sociologie des controverses technologiques » dans Madeleine Akrich, Michel Callon et Bruno Latour (eds.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'Ecole des Mines, 2006.

CALLON Michel, « Sociologie de l'acteur réseau » dans Madeleine Akrich, Michel Callon et Bruno Latour (eds.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'Ecole des Mines, 2006, p. 267-276.

CALLON Michel, « Ni intellectuel engagé, ni intellectuel dégagé : la double stratégie de l'attachement et du détachement », *Sociologie du travail*, 1999, n° 41, p. 65-78.

CALLON Michel, « La sociologie peut-elle enrichir l'analyse économique des externalités ? Essai sur la notion de cadrage-débordement » dans Dominique Foray et Jacques Mairesse (eds.), *Innovations et performances. Approches interdisciplinaires*, Paris, Editions de l'EHESS, 1999, p. 399-431.

CALLON Michel, « Éléments pour une sociologie de la traduction. La domestication des coquilles Saint-Jacques dans la Baie de Saint-Brieuc », *L'Année sociologique*, 1986, n° 36.

CALLON Michel et LATOUR Bruno, « Le grand Léviathan s'apprivoise-t-il ? » dans Madeleine Akrich, Michel Callon et Bruno Latour (eds.), *Sociologie de la traduction. Textes fondateurs*, Paris, Presses de l'Ecole des Mines, 2006, p. 11-32.

CALLON Michel et LATOUR Bruno (eds.), *La science telle qu'elle se fait. Anthologie de la sociologie des sciences de langue anglaise*, Paris, La Découverte, 1990, 396 p.

CALLON Michel, LAW John et RIP Arie (eds.), *Mapping the Dynamics of Science and Technology : Sociology of Science in the Real World*, London, Mac Millan, 1986, 242 p.

CALLON Michel, MILLO Yuval et MUNIESA Fabian, *Market Devices*, Malden, Blackwell Publishing, 2007.

CARDON Dominique, *A quoi rêvent les algorithmes. Nos vies à l'heure des big data*, Paris, Seuil, 2015, 112 p.

CARDON Dominique et LEVREL Julien, « La vigilance participative. Une interprétation de la gouvernance de Wikipédia », *Réseaux*, 2009, vol. 2, n° 154, p. 51-89.

CARR Nicholas, *Google in the middle*, <http://www.routhtype.com/?p=1249> , 10 avril 2009, consulté le 15 janvier 2015.

CEFAÏ Daniel, « Introduction. L'engagement ethnographique. » dans Daniel Céfaï (ed.), *L'engagement ethnographique*, Paris, Editions de l'EHESS, 2010, p. 7-21.

CEFAÏ Daniel, « Qu'est-ce qu'une arène publique ? Quelques pistes pour une approche pragmatiste » dans Daniel Cefaï et Isaac Joseph (eds.), *L'Héritage du pragmatisme. Conflits d'urbanité et épreuves de civisme*, La Tour d'Aigues, Editions de l'Aube, 2002, p. 51-82.

CEFAÏ Daniel, « La construction des problèmes publics. Définitions de situations dans des arènes publiques. », *Réseaux*, 1996, vol. 14, n° 75, p. 43-66.

CHARVOLIN Florian, *L'invention de l'environnement en France*, Paris, La Découverte, 2003, 133 p.

CHAVEZ-DREYFUSS Gertrude, « Honduras to build land title registry using bitcoin technology », *Reuters*, 15 mai 2015.

CHIAPELLO Eve, « Accounting and the Birth of the Notion of Capitalism », *Critical Perspective on Accounting*, 2007, n° 18, p. 263-296.

CHRISMAN Nicholas, *Exploring geographic information systems*, 2nd éd., New York, Wiley, 2002, 320 p.

CLARKE Amanda et MARGETTS Helen, « Governments and Citizens Getting to Know Each Other? Open, Closed, and Big Data in Public Management Reform », *Policy & Internet*, 2014, vol. 6, n° 4, p. 393-417.

COLIN Nicolas et VERDIER Henri, *L'âge de la multitude. Entreprendre et gouverner après la révolution numérique.*, Paris, Armand Colin, 2012.

COLLIN Pierre et COLIN Nicolas, *Mission d'expertise sur la fiscalité de l'économie numérique*, Paris, Ministère de l'économie et des finances, 2013.

CONRADIE Peter et CHOENNI Sunil, « On the barriers for local government releasing open data », *Government Information Quarterly*, 2014, vol. 31, p. S10-S17.

CRAMPTON Jeremy, « Cartography: Maps 2.0 », *Progress in Human Geography*, 2009, vol. 33, n° 1, p. 91-100.

DAGIRAL Eric et PEERBAYE Ashveen, « Les mains dans les bases de données. Connaître et faire reconnaître le travail invisible », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 191-216.

DAVID Paul A. et BUNN Julie Ann, « The economics of gateway technologies and network evolution: Lessons from electricity supply history », *Information Economics and Policy*, 1988, vol. 3, n° 2, p. 165-202.

DEBARBIEUX Bernard, « Territoire-Territorialité-Territorialisation : aujourd'hui encore, et bien moins que demain... » dans Martin Vanier (ed.), *Territoires, Territorialité, Territorialisation. Controverses et perspectives*, Rennes, Presses Universitaires de Rennes, 2009, p. 75-89.

DE FILIPPI Primavera et MAURO Raffaele, « Ethereum: the decentralised platform that might displace today's institutions », *Internet Policy Review*, 25 août 2014.

DENEGRE Jean et SALGE François, *Les systèmes d'information géographique, Que sais-je ?*, Paris, Presses Universitaires de France - PUF, 2004, 128 p.

DENIS Jérôme, « L'informatique et sa sécurité. Le souci de la fragilité technique. », *Réseaux*, 2012, vol. 30, n° 171, p. 161-187.

DENIS Jérôme et GOËTA Samuel, « La fabrique des données brutes. Le travail en coulisses de l'open data. » dans Clément Mabi, Jean-Christophe Plantin et Laurence Monnoyer-Smith (eds.), *Penser l'écosystème des données. Les enjeux scientifiques et politiques des données numériques.*, Paris, Editions de la Maison des Sciences de l'Homme, 2015.

DENIS Jérôme et PONTILLE David, « Une infrastructure évasive. Aménagements cyclables et troubles de la description dans OpenStreetMap », *Réseaux*, 2013, vol. 2, n° 178-179, p. 91-125.

DENIS Jérôme et PONTILLE David, « Travailleurs de l'écrit, matières de l'information », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 1-20.

DESAGE Fabien et GUERANGER David, *La politique confisquée : Sociologie des réformes et des institutions intercommunales*, Bellecombe-en-Bauges, Editions du Croquant, 2011, 247 p.

DESROSIERES Alain, *Prouver et gouverner. Une analyse politique des statistiques publiques.*, Paris, La Découverte, 2014, 283 p.

DESROSIERES Alain, *Pour une sociologie historique de la quantification*, Paris, Presses de l'École des mines, 2008, 328 p.

DESROSIERES Alain, « La statistique, outil de gouvernement et outil de preuve » dans *Pour une sociologie historique de la quantification*, Paris, Presses de l'École des Mines, 2008, p. 7-20.

DESROSIERES Alain, « Décrire l'Etat ou explorer la société : les deux sources de la statistique publique », *Genèses*, 2005, n° 58, p. 4-27.

DESROSIERES Alain, « Historiciser l'action publique. L'État, le marché et les statistiques. » dans Pascale Laborier et Danny Trom (eds.), *Historicités de l'action publique*, Paris, PUF, 2003.

DESROSIERES Alain, « Les qualités des quantités », *Courrier des statistiques*, 2003, n° 105-106, p. 51-63.

DESROSIERES Alain, « Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative. », *Genèses*, 2001, vol. 2, n° 43, p. 112-127.

DESROSIERES Alain, « Le territoire et la localité. Deux langages statistiques. », *Politix*, 1994, n° 25, p. 46-58.

DESROSIERES Alain, *La politique des grands nombres. Histoire de la raison statistique.*, Paris, La Découverte, 1993.

DESROSIERES Alain, « Discuter l'indiscutable. Raison statistique et espace public. », *Raisons Pratiques*, 1992, n° 3, p. 131-154.

DEWEY John, *Le public et ses problèmes*, Paris, Gallimard, 2010.

DIDIER Emmanuel, « La consistance du futur. La prévision statistique aux États-Unis et la Grande Dépression », *Raisons politiques*, 2012, vol. 4, n° 48, p. 65-83.

DIDIER Emmanuel, *En quoi consiste l'Amérique ? Les statistiques, le New Deal et la démocratie*, Paris, La Découverte, 2009, 320 p.

DIDIER Emmanuel, NEVANEN Sophie, ROBERT Philippe et ZAUBERMAN Renée, « La solidité des institutions. Les statistiques de “victimation” de l’Insee (1996-2006) », *Genèses*, 2009, vol. 1, n° 74, p. 128-144.

DONOVAN Kevin P., « Seeing Like a Slum: Towards Open, Deliberative Development », *Georgetown Journal of Informational Affairs*, 2012, vol. 13, n° 1, p. 97-104.

DOWEK Gilles, *Introduction à la science informatique*, Paris, Centre régional de documentation pédagogique de l’académie de Paris, 2011, 375 p.

DOWNEY Paul, *Registers: authoritative lists you can trust*, <https://gds.blog.gov.uk/2015/09/01/registers-authoritative-lists-you-can-trust/>, 9 janvier 2015, consulté le 15 décembre 2015.

DUCHESNE Sophie et HAEGEL Florence (eds.), « Repérages du politique. Regards disciplinaires et approches de terrain », *Espaces Temps*, 2001, vol. 76, n° 1, p. 1-154.

DURAN Patrice, « Territorialisation » dans Romain Pasquier, Sébastien Guigner et Alistair Cole (eds.), *Dictionnaire des politiques territoriales*, Paris, Presses de Sciences Po, 2011.

DURAN Patrice, « Légitimité, droit et action publique », *L’Année sociologique*, 2009, vol. 59, n° 2, p. 303-344.

DURAN Patrice, « Action publique et pouvoir politique. La difficile conciliation de la légitimité et de l’efficacité » dans *Penser l’action publique*, Paris, LGDJ, 1999, p. 75-112.

EDWARDS Paul, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.*, Cambridge Mass., MIT Press, 2010.

EDWARDS Paul, JACKSON Steven, BOWKER Geoffrey C. et WILLIAMS Robin, « Introduction: An Agenda for Infrastructure Studies », *Journal of the Association for Information Systems*, 2009, vol. 10, n° 5, p. 364-374.

ELDEN Stuart, *The Birth of Territory*, Chicago, University of Chicago Press, 2013.

ELDEN Stuart, « Land, terrain, territory », *Progress in Human Geography*, 2010, vol. 34, n° 6, p. 799-817.

ELDEN Stuart, « Governmentality, Calculation, Territory », *Environment and Planning D: Society and Space*, 2007, vol. 25, n° 3, p. 562-580.

ELWOOD Sarah, « Volunteered geographic information: future research directions motivated by critical, participatory, and feminist GIS », *GeoJournal*, 2008, vol. 72, n° 3, p. 173-183.

EPSTEIN Renaud, « La gouvernance territoriale : une affaire d'État La dimension verticale de la construction de l'action collective dans les territoires. », *L'Année sociologique*, 2015, vol. 65, n° 2, p. 457-482.

EPSTEIN Renaud, « Gouverner à distance. Quand l'Etat se retire des territoires. », *Esprit*, 2006, n° 11, p. 96-111.

EROS Emily, MEHNDIRATTA Shomik, ZEGRAS Chris, WEBB Kevin et OCHOA Maria C, « Applying the General Transit feed Specification to the Global South: Experiences in Mexico City, Mexico, and Beyond », *Transportation Research Record: Journal of the Transportation Research Board*, 2014, p. 44-52.

ESPELAND Wendy Nelson et STEVENS Mitchell L., « A Sociology of Quantification », *European Journal of Sociology*, 2008, vol. 49, n° 03, p. 401-436.

ESPELAND Wendy Nelson et STEVENS Mitchell L., « Commensuration as a Social Process », *Annual Review of Sociology*, 1998, vol. 24, p. 313-343.

FANIEL Ixchel M. et JACOBSEN Trond E., « Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data », *Computer Supported Cooperative Work*, 2010, vol. 19, n° 3, p. 355-375.

FAUCONNET Paul, *La responsabilité : étude de sociologie*, Paris, Félix Alcan, 1920, 400 p.

FAURE Alain et DOUILLET Anne-Cécile, *L'action publique et la question territoriale*, Grenoble, Presses Universitaires de Grenoble, 2005.

FAURE Alain, LERESCHE Jean-Philippe, MULLER Pierre et NAHRATH Stéphane (eds.), *Action publique et changements d'échelles : les nouvelles focales du politique*, Paris, L'Harmattan (coll. « Logiques politiques »), 2007, 380 p.

FAVRE Pierre, « Ce que les science studies font à la science politique. Réponse à Bruno Latour. », *Revue française de science politique*, 2008, vol. 58, n° 5, p. 817-829.

FAVRE Pierre, « La question de l'objet de la science politique a-t-elle un sens ? » dans Pierre Favre, Olivier Fillieule et Fabien Jobard (eds.), *L'atelier du politiste. Théories, actions, représentations.*, Paris, La Découverte, 2007, p. 17-33.

FEYT Grégoire, « Les métiers du territoire face aux technologies de l'information géographique : Babel et esperanto » dans Stéphane Roche et Claude Carron (eds.), *Aspects organisationnels des SIG*, Paris, Hermès, 2004, p. 55-70.

FEYT Grégoire et LARDON Sylvie, « L'émergence de nouveaux référentiels pour dire le territoire : vers une complémentarité des points de vue ou un "combat des cartes" ? », Montpellier, 2010.

FLICHY Patrice, « Rendre visible l'information. Une analyse sociotechnique du traitement des données. », *Réseaux*, février 2013, n° 178-179, p. 55-89.

FLONNEAU Mathieu et GUIGUENO Vincent (eds.), *De l'histoire des transports à l'histoire de la mobilité ?*, Rennes, Presses Universitaires de Rennes, 2009, 334 p.

FLOWERS Michael, « Beyond Open Data: The Data-Driven City » dans Brett Goldstein et Lauren Dyson (eds.), *Beyond Transparency*, San Francisco, Code for America Press, 2013, p. 185-199.

FOLI Olivia et DULAURANS Marlène, « Tenir le cap épistémologique en thèse Cifre. Ajustements nécessaires et connaissances produites en contexte », *Etudes de communication*, 2013, n° 40, p. 59-76.

FOUCAULT Michel, *Sécurité territoire population*, Paris, Seuil, 2004, 448 p.

FOUILLE Laurent, *L'attachement automobile mis à l'épreuve. Etude des dispositifs de détachement et de recomposition des mobilités.*, Université Rennes 2, Rennes, 2010, 476 p.

FRENOT Stéphane et GRUMBACH Stéphane, « Des données à l'intermédiation, une révolution économique et politique. » dans Lisette Calderan et al. (eds.), *Big Data : nouvelles partitions de l'information*, Bruxelles, De Boeck, 2015, p. 97-120.

GAFFNEY Christopher et ROBERTSON Cerianne, « Smarter than Smart: Rio de Janeiro's Flawed Emergence as a Smart City », *Journal of Urban Technology*, 2016, p. 1-18.

GAGLIO Gérald, « En quoi une thèse CIFRE en sociologie forme au métier de sociologue ? Une hypothèse pour ouvrir le débat », *Socio-logos*, 2008, n° 3.

GALIMBERTI Deborah, LOBRY Sylvaine, PINSON Gilles et RIO Nicolas, « La métropole de Lyon. Splendeurs et fragilités d'une machine intercommunale. », *Hérodote*, 2014, vol. 3, n° 154, p. 191-209.

GARDEY Delphine, *Le Linge du Palais-Bourbon. Corps, matérialité et genre du politique à l'ère démocratique*, Lormont, Le Bord de l'Eau, 2015, 256 p.

GARDEY Delphine, *Ecrire, calculer, classer : Comment une révolution de papier a transformé les sociétés contemporaines*, Paris, La Découverte, 2008, 320 p.

GAUDIN Jean-Pierre, « Contrats » dans Laurie Boussaguet, Sophie Jacquot et Pauline Ravinet (eds.), *Dictionnaire des politiques publiques*, Paris, Presses de Sciences Po, 2010, p. 164-171.

GAUDIN Jean-Pierre, *Gouverner par contrat*, Paris, Presses de Sciences Po, 2007, 280 p.

GEERTZ Clifford, « Du point de vue de l'indigène : sur la nature de la compréhension anthropologique. » dans *Savoir local, savoir global. Les lieux du savoir.*, Paris, PUF, 1986.

GEHL Robert, « Sharing, knowledge management and big data: A partial genealogy of the data scientist », *European Journal of Cultural Studies*, 2015, vol. 18, n° 4-5, p. 413-428.

GERSON Elihu M. et STAR Susan Leigh, « Analyzing due process in the workplace », *ACM Transactions on Office Information Systems*, 1986, vol. 4, n° 3, p. 257-270.

GILBERT Claude, HENRY Emmanuel et BOURDEAUX Isabelle, « Lire l'action publique au prisme des processus de définition des problèmes » dans Claude Gilbert et Emmanuel Henry (eds.), *Comment se construisent les problèmes de santé publique*, Paris, La Découverte, 2009, p. 7-33.

GILLEPSIE Tarleton, « The politics of "platforms" », *New Media & Society*, 2010, vol. 12, n° 3, p. 347-364.

GIRARD Anne et ROUSSEL Violaine, « Une question de confiance. », *Raisons Pratiques*, 2003, vol. 2, n° 10, p. 171-186.

GITELMAN Lisa (ed.), « *Raw Data* » *Is an Oxymoron*, Cambridge Mass., MIT Press, 2013, 192 p.

GOËTA Samuel, « The Daily Shaping of State Transparency: Emerging Standards in Open Government Data », Buenos Aires, 2014.

GOËTA Samuel, *Open data : qu'ouvre-t-on avec les données publiques ?*, Université de Paris IV - Sorbonne, Paris, 2011, 101 p.

GOËTA Samuel, « Des données à la recherche de publics: les assemblages temporaires des concours de réutilisation de données ouvertes », *Participations*, à paraître.

GOFFMAN Erving, *Les cadres de l'expérience*, Paris, Les Editions de Minuit, 1991, 576 p.

GOMART Emilie, « Methadone: Six effects in Search of a Substance », *Social Studies of Science*, 2002, vol. 32, n° 1, p. 93-135.

GOMART Emilie, *Surprised by Methadone*, Ecole des Mines, Paris, 1999.

GOMART Emilie et HENNION Antoine, « A Sociology of Attachment: Music Lovers, Drug Addicts » dans John Law et John Hassard (eds.), *Actor Network Theory and After*, Oxford/Malden MA, Blackwell Publishers, 1999.

GOODCHILD Michael, « Citizens as sensors: the world of volunteered geography », *GeoJournal*, 2007, vol. 69, n° 4, p. 211-221.

GOULET Frédéric et LE VELLY Ronan, « Comment vendre un produit incertain ? Activités de détachement et d'attachement d'une firme d'agrofourmiture », *Sociologie du Travail*, 2013, vol. 55, n° 3, p. 369-386.

GOULET Frédéric et VINCK Dominique, « L'innovation par retrait. Contribution à une sociologie du détachement. », *Revue française de sociologie*, 2012, vol. 53, n° 2, p. 195-224.

GRAY Jonathan, « Towards a Genealogy of Open Data », Glasgow, 2014.

Fighting Phantom Firms in the UK: From Opening Up Datasets to Reshaping Data Infrastructures?, 2015.

GREENFIELD Adam, *Against the Smart City*, London, Do Projects, 2013.

GURSTEIN Michael, « Open data: Empowering the empowered or effective data use for everyone? », *First Monday*, 2011, vol. 16, n° 2.

GUSFIELD Joseph, *La culture des problèmes publics : L'alcool au volant : la production d'un ordre symbolique*, Paris, Economica, 2008, 354 p.

HACKING Ian, « Façonner les gens : le seuil de pauvreté » dans Jean-Pierre Beaud et Jean-Guy Prévost (eds.), *L'ère du chiffre. Systèmes statistiques et traditions nationales*, Montréal, Presses de l'université du Québec, 2000, p. 17-36.

HALPERN Charlotte et LE GALES Patrick, « Pas d'action publique autonome sans instruments propres. Analyse comparée et longitudinale des politiques environnementales et urbaines de l'Union européenne », *Revue française de science politique*, 2011, vol. 61, n° 1, p. 51-78.

HARAWAY Donna, *When Species Meet*, Minneapolis, University of Minnesota Press, 2008, 423 p.

HARMAN Graham, *Bruno Latour: Reassembling the Political*, London, Pluto Press, 2014, 216 p.

HARVEY Francis et CHRISMAN Nick, « Boundary objects and the social construction of GIS technology », *Environment and Planning A*, 1998, vol. 30, n° 9, p. 1683-1694.

HELBIG Natalie, CRESSWELL Anthony M., BURKE Brian et LUNA-REYES Luis, « The dynamics of opening government data », *Center for Technology in Government*, 2012, p. 34.

HENNION Antoine, « D'une sociologie de la médiation à une pragmatique des attachements. », *SociologieS*, 2013.

HENNION Antoine, *La passion musicale. Une sociologie de la médiation*, Paris, Métailié, 1993.

HILL Dan, « On the smart city; Or, a “manifesto” for smart citizens instead », *City of Sound*, 1 février 2013.

HOLLANDS Robert G., « Critical interventions into the corporate smart city », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 61-77.

HOLLANDS Robert G., « Will the real smart city please stand up? », *City*, 2008, vol. 12, n° 3, p. 303-320.

HOOD Christopher, *The Tools of Government*, London, Macmillan, 1983, 178 p.

HOPWOOD Anthony et MILLER Peter (eds.), *Accounting as Social and Institutional Practice*, Cambridge, Cambridge University Press, 1994, 340 p.

HOWARD Alex, *Predictive data analysis is saving lives and taxpayer dollars*, <http://radar.oreilly.com/2012/06/predictive-data-analytics-big-data-nyc.html> , 26 juin 2012, consulté le 23 octobre 2015.

JANSSEN Katleen, *The Availability of Spatial and Environmental Data in the EU. At the Crossroads between Public and Economic Interests*, Alphen a/d Rijn, Kluwer Law International, 2010, 656 p.

JANSSEN Marijn, CHARALABIDIS Yannis et ZUIDERWIJK Anneke, « Benefits, Adoption Barriers and Myths of Open Data and Open Government », *Information Systems Management (ISM)*, vol. 29, n° 4, p. 258-268.

JOLIVEAU Thierry, « Gérer l'environnement avec des SIG. Mais qu'est-ce qu'un SIG ? », *Revue de géographie de Lyon*, 1996, vol. 71, n° 2, p. 101-110.

JOUBE Bernard, *Les politiques de déplacements urbains en Europe. L'innovation en question dans cinq villes européennes.*, Paris, L'Harmattan, 2003, 191 p.

JOUBE Bernard et LEFEVRE Christian, « De la gouvernance urbaine au gouvernement des villes ? Permanence ou recomposition des cadres de l'action publique en Europe », *Revue française de science politique*, 1999, vol. 49, n° 6, p. 835-853.

KEEN Justin, CALINESCU Radu, PAIGE Richard et ROOKSBY John, « Big data + politics = open data: The case of health care data in England », *Policy & Internet*, 2013, vol. 5, n° 2, p. 228-243.

KITCHIN Rob, *The data revolution: Big data, open data, data infrastructures and their consequences*, London, SAGE publications, 2014.

KITCHIN Rob, *Four critiques of open data initiatives*, <http://blogs.lse.ac.uk/impactofsocialsciences/2013/11/27/four-critiques-of-open-data-initiatives/>, 27 novembre 2013.

KITCHIN Rob et MCARDLE Gavin, « The diverse nature of big data », *Big Data & Society*, 2016, vol. 3, n° 1, p. 1-10.

KLAUSER Francisco, PAASCHE Till et SÖDERSTRÖM Ola, « Michel Foucault and the Smart City: Power Dynamics Inherent in Contemporary Governing through Code », *Environment and Planning D: Society and Space*, 2014, vol. 32, p. 869-885.

KOOIMAN Jan, *Governing as Governance*, London, Sage Publications, 2003, 256 p.

LABELLE Sarah et LE CORF Jean-Baptiste, « Modalités de diffusion et processus documentaires, conditions du "détachement" des informations publiques. Analyse des discours législatifs et des portails open data territoriaux. », *Les enjeux de l'information et de la communication*, 2012, n° 2012/1, p. 209-222.

LABORIER Pascale, « Légitimité » dans Laurie Boussaguet, Sophie Jacquot et Pauline Ravinet (eds.), *Dictionnaire des politiques publiques*, 3e édition., Paris, Presses de Sciences Po, 2010, p. 335-343.

LAGROYE Jacques, *La politisation*, Paris, Belin, 2003, 576 p.

LAGROYE Jacques, « La légitimation » dans Madeleine Grawitz et Jean Leca (eds.), *Traité de science politique*, Paris, Presses Universitaires de France, 1985, p. 395-467.

LANDOUR Julie, « Le chercheur funambule. Quand une salariée se fait la sociologue de son univers professionnel », *Genèses*, 2013, vol. 1, n° 90, p. 25-41.

LANEY Doug, *3D Data Management: Controlling Data Volume, Velocity, and Variety*, Stamford, META Group, 2001.

LASCOUMES Pierre, *Action publique et environnement*, Que sais-je ?, Paris, Presses Universitaires de France, 2012, 128 p.

LASCOUMES Pierre, « La Gouvernamentalité : de la critique de l'État aux technologies du pouvoir », *Le Portique*, 2004, n° 13-14.

LASCOUMES Pierre, « Rendre gouvernable : de la "traduction" au "transcodage". L'analyse des processus de changement dans les réseaux d'action publique. » dans CURAPP (ed.), *La gouvernabilité*, PUF., Paris, 1996.

LASCOUMES Pierre, *L'éco-pouvoir. Environnements et politiques.*, Paris, La Découverte, 1994, 324 p.

LASCOUMES Pierre et LE GALES Patrick, *Sociologie de l'action publique*, Paris, Armand Colin, 2007.

LASCOUMES Pierre et LE GALES Patrick, « Introduction : L'action publique saisie par ses instruments » dans *Gouverner par les instruments*, Paris, Presses de Sciences Po, 2005.

LASCOUMES Pierre et LE GALES Patrick (eds.), *Gouverner par les instruments*, Paris, Presses de Sciences Po, 2004, 369 p.

LASCOUMES Pierre et SIMARD Louis, « L'action publique au prisme de ses instruments. », *Revue française de science politique*, 2011, vol. 61, n° 1, p. 5-22.

LATOUR Bruno, « An Attempt at a "Compositionist Manifesto" », *New Literary History*, 2010, vol. 41, n° 3, p. 471-490.

LATOUR Bruno, « Pour un dialogue entre science politique et science studies », *Revue française de science politique*, 2008, vol. 58, n° 4, p. 657-678.

LATOUR Bruno, « Paris, ville invisible : le plasma » dans Christine Macel, Daniel Birnbaum et Valérie Guillaume (eds.), *Airs de Paris, 30 ans du Centre Pompidou*, Paris, ADGP, 2007, p. 260-263.

LATOUR Bruno, *La science en action : Introduction à la sociologie des sciences*, 3e éd., Paris, La Découverte, 2005.

LATOUR Bruno, *Reassembling the Social. An introduction to Actor-Network Theory*, Oxford, Oxford University Press, 2005.

LATOUR Bruno, « De la politique du temps à la politique de l'espace ? » dans Elisabeth Rémy et al. (eds.), *Espaces, savoirs et incertitudes*, Paris, Ibis Press, 2005, p. 5-12.

LATOUR Bruno, « Factures/Fractures. De la notion de réseau à celle d'attachement » dans André Micoud et Michel Peroni (eds.), *Ce qui nous relie*, La Tour d'Aigues, Editions de l'Aube, 2000, p. 189-208.

LATOUR Bruno, *Politiques de la nature. Comment faire entrer les sciences en démocratie*, Paris, La Découverte, 1999, 383 p.

LATOUR Bruno, « Le topofil de Boa Vista ou la référence scientifique -montage photo-philosophique », *Raison Pratique*, 1993, n° 4, p. 187-216.

LATOUR Bruno, « Le “topofil” de Boa Vista ou la référence scientifique » dans *La clef de Berlin et autres leçons d'un amateur de science*, Paris, La Découverte, 1993.

LATOUR Bruno, *Aramis ou l'amour des techniques*, Paris, La Découverte, 1992, 248 p.

LATOUR Bruno, « Les “vues” de l'esprit », *Culture technique*, 1985, n° 14, p. 4-30.

LATOUR Bruno, *Pasteur : guerre et paix des microbes : Suivi de Irréductions*, Paris, La Découverte, 1984, 363 p.

LATOUR Bruno et HERMANT, *Paris, ville invisible, Les empêcheurs de penser en rond.*, Paris, La Découverte, 1998, 159 p.

LATOUR Bruno, JENSEN Pablo, VENTURINI Tommaso, GRAUWIN Sébastien et BOULLIER Dominique, « 'The whole is always smaller than its parts' – a digital test of Gabriel Tarde's monads », *The British Journal of Sociology*, 2012, vol. 63, n° 4, p. 590-615.

LEFEBVRE Henri, *Le droit à la ville*, 3e édition., Paris, Economica, 2009, 135 p.

LE GALES Patrick, *Le retour des villes européennes : Sociétés urbaines, mondialisation, gouvernement et gouvernance*, 2e édition., Paris, Presses de Sciences Po, 2011.

LE GALES Patrick, « Gouvernance » dans Laurie Boussaguet, Sophie Jacquot et Pauline Ravinet (eds.), *Dictionnaire des politiques publiques*, Paris, Presses de Sciences Po, 2010.

LE GALES Patrick, « Du gouvernement des villes à la gouvernance urbaine », *Revue française de science politique*, 1995, vol. 45, n° 1, p. 57-95.

LELONG Benoît et MALLARD Alexandre, « La fabrication des normes », *Réseaux*, 2000, vol. 18, n° 102, p. 9-34.

LEMAIRE Elodie, « “En être” et “s’en défaire”. Retour réflexif sur une enquête par immersion dans la police », Rennes, 2013.

LEMIEUX Cyril, « Peut-on ne pas être constructiviste ? », *Politix*, 2012, vol. 4, n° 100, p. 169-187.

LEMIEUX Cyril, « À quoi sert l’analyse des controverses ? », *Mil neuf cent*, janvier 2007, n° 25, p. 191-212.

LEMOINE Benjamin, *Les valeurs de la dette. L’Etat à l’épreuve de la dette publique.*, Ecole nationale supérieure des mines de Paris, Paris, 2011, 426 p.

LE TILLY Yann, *Google Transit à Paris : pour les cowboys uniquement !*, <http://transid.blogspot.fr/2012/11/google-transit-paris-pour-les-cowboys.html> , 28 novembre 2012, consulté le 24 octobre 2015.

LEVY Jacques et LUSSAULT Michel (eds.), *Dictionnaire de la géographie et de l’espace des sociétés*, Paris, Belin, 2003.

LIPPMANN Walter, *The phantom public*, New York, Simon & Schuster, 1927.

LONGO Justin, « Open Data: Digital-Era Governance Thoroughbred or New Public Management Trojan Horse? », *Public Policy and Governance Review*, 2011, vol. 2, n° 2, p. 38-51.

LORRAIN Dominique et LE GALES Patrick, « Gouverner les très grandes métropoles », *Revue française d’administration publique*, 2003, vol. 3, n° 107, p. 305-317.

LOVELUCK Benjamin, *Réseaux, libertés et contrôle. Une généalogie politique d'internet*, Paris, Armand Colin, 2015.

LOVELUCK Benjamin, « Internet, une société contre l'Etat ? Libéralisme informationnel et économies politiques de l'auto-organisation en régime numérique », *Réseaux*, 2015, vol. 4, n° 192, p. 235-270.

LUHMANN Niklas, *La confiance, un mécanisme de réduction de la complexité sociale*, Paris, Economica, 2006, 123 p.

LUSSAULT Michel, *L'Homme spatial : La construction sociale de l'espace humain*, Paris, Seuil, 2007, 363 p.

LUSSAULT Michel, « Territorial (Modèle) » dans Jacques Lévy et Michel Lussault (eds.), *Dictionnaire de la géographie et de l'espace des sociétés*, Paris, Belin, 2003, p. 917-918.

MARTIN Chris, « Barriers to the Open Government Data Agenda: Taking a Multi-Level Perspective », *Policy & Internet*, 2014, vol. 6, n° 3.

MARTIN Olivier, *La Mesure de l'esprit. Origines et développement de la psychométrie, 1900-1950*, Paris, L'Harmattan, 1997, 384 p.

MAYER-SCHÖNBERGER Viktor et CUKIER Kenneth, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, New York, Houghton Mifflin Harcourt, 2013, 240 p.

MCFALL Liz, « What Have Market Devices Got to Do with Public Policy? For Regulation, Market Devices and Policy Instruments Collection » dans Charlotte Halpern, Pierre Lascombes et Patrick Le Galès (eds.), *L'instrumentation de l'action publique.*, Paris, Presses de Sciences Po, 2014, p. 119-142.

MCHUGH Bibiana, « Pioneering open data standards: The GTFS story » dans Brett Goldstein et Lauren Dyson (eds.), *Beyond Transparency*, San Francisco, Code for America Press, 2013, p. 125-125.

MCNEILL Donald, « Global firms and smart technologies: IBM and the reduction of cities », *Transactions of the Institute of British Geographers*, 2015, vol. 40, n° 4, p. 562-574.

MEIJER Albert et BOLIVAR Manuel Pedro Rodriguez, « Governing the smart city: a review of the literature on smart urban governance », *International Review of Administrative Sciences*, 2016, vol. 82, n° 2, p. 392-408.

MILLERAND Florence, « La science en réseau. Les gestionnaires d'information « invisibles »
Antoine COURMONT – « Politiques des données urbaines » - Thèse IEP de Paris – 2016

dans la production d'une base de données scientifiques », *Revue d'anthropologie des connaissances*, 2012, vol. 6, n° 1, p. 163-190.

MOL Annemarie, *The Body Multiple: Ontology in Medical Practice.*, London, Duke University Press, 2003, 216 p.

MOROZOV Evgeny, *Le mirage numérique : Pour une politique du Big Data*, Paris, Les prairies ordinaires, 2015, 130 p.

MOROZOV Evgeny, *To Save Everything, Click Here: The Folly of Technological Solutionism*, New York, Public Affairs, 2013, 432 p.

NACHI Mohamed, *Introduction à la sociologie pragmatique*, Paris, Armand Colin, 2006.

NAKAMOTO Satoshi, « Bitcoin: A Peer-to-Peer Electronic Cash System », 2008.

NEFF Gina et STARK David, « Permanently Beta: Responsive Organization in the Internet Era » dans Philip N. Howard et Steve Jones (eds.), *Society Online. The Internet in Context*, Thousand Oaks, Sage Publications, 2004, p. 173-188.

NIVIERE Delphine, « Négocier une statistique européenne : le cas de la pauvreté », *Genèses*, 2005, vol. 1, n° 58, p. 28-47.

NOUCHER Matthieu, « Infrastructures de données géographiques et flux d'information environnementale », *Netcom*, 2013, vol. 27, n° 1-2, p. 120-147.

OLIVIER DE SARDAN Jean-Pierre, « La politique du terrain. Sur la production des données en anthropologie. », *Enquête*, 1995, n° 1, p. 71-109.

O'REILLY Tim, « Government as a Platform », *innovations*, 2010, vol. 6, n° 1, p. 13-40.

OSTROM Elinor, *La gouvernance des biens communs. Pour une nouvelle approche des ressources naturelles.*, Bruxelles, De Boeck, 2010, 301 p.

PADIOLEAU Jean-Gustave, « L'action publique urbaine moderniste », *Politiques et Management Public*, 1991, vol. 9, n° 3, p. 133-146.

PARK Robert Ezra, « La ville. Propositions de recherche sur le comportement humain en milieu urbain » dans Isaac Joseph et Yves Grafmeyer (eds.), *L'école de Chicago : Naissance de l'écologie urbaine*, Paris, Flammarion, 2005, p. 83-125.

PASQUIER Romain, SIMOULIN Vincent et WEISBEIN Julien (eds.), *La gouvernance*
Antoine COURMONT – « Politiques des données urbaines » - Thèse IEP de Paris – 2016

territoriale : Pratiques, discours et théories, Paris, LGDJ, 2007, 235 p.

PASSERON Jean-Claude et REVEL Jacques, *Penser par cas*, Paris, Editions de l'EHESS, 2005, 292 p.

PENIN Julien, HUSSLER Caroline, MILLOT Valentine, RONDE Patrick et SEYED-RASOLI Haniyeh, *La valorisation des informations du secteur public (ISP) : un modèle économique de tarification optimale*, Strasbourg, BETA (Bureau d'Economie Théorique et Appliquée), 2010.

PERKINS Chris, « Plotting practices and politics: (im)mutable narratives in OpenStreetMap », *Transactions of the Institute of British Geographers*, 2014, vol. 39, n° 2, p. 304-317.

PESTRE Dominique, *Introduction aux Science Studies*, Repères., Paris, La Découverte, 2006, 128 p.

PICON Antoine, *Smart Cities. Théorie et critique d'un idéal auto-réalisateur.*, Paris, Editions B2, 2013, 117 p.

PIJOURLET Pierre, « Le système urbain de références de la communauté urbaine de Lyon » dans *Forma Urbis - Les plans généraux de Lyon : XVIe-XXe siècle*, Lyon, Archives municipales de Lyon, 1997.

PINSON Gilles, « La gouvernance des villes françaises. Du schéma centre-périphérie aux régimes urbains », *Pôle Sud*, 2010, vol. 1, n° 32, p. 73-92.

PORTER Theodore, *Trust in numbers: the pursuit of objectivity in science and public life*, Princeton, Princeton University Press, 1995, 326 p.

PORTER Theodore, « Making things quantitative », *Science in context*, 1994, vol. 7, n° 3, p. 389-407.

RABARI Chirag et STORPER Michael, « The digital skin of cities: urban theory and research in the age of the sensed and metered city, ubiquitous computing and big data », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 27-42.

RAFFESTIN Claude, « Espace et pouvoir » dans A Bailly (ed.), *Les concepts de la géographie humaine*, Paris, Masson, 1995.

RIFFAULT Tristan, « Fabriquer de la politique. La production du discours politique lors d'une campagne électorale (2013-2014) », *Réseaux*, 2015, vol. 2, n° 190-191, p. 13-44.

RIFKIN Jeremy, *L'âge de l'accès*, Paris, La Découverte, 2005, 406 p.

ROBINSON David, YU Harlan, ZELLER William et FELTEN Edward, « Government Data and the Invisible Hand », *Yale Journal of Law & Technology*, 2009, vol. 11, n° 1.

ROCHE Stéphane, *Les enjeux sociaux des systèmes d'information géographique*, Paris, L'Harmattan, 2000, 158 p.

ROSENBERG Daniel, « Data before the fact » dans Lisa Gitelman (ed.), *'Raw Data' is an Oxymoron*, Cambridge Mass., MIT Press, 2013, p. 15-40.

ROSSI Ugo, « The Variegated Economics and the Potential Politics of the Smart City », *Territory, Politics, Governance*, 2016, vol. 4, n° 3, p. 337-353.

ROTH Matthew, *How Google and Portland's TriMet Set the Standard for Open Transit Data*, <http://sf.streetsblog.org/2010/01/05/how-google-and-portlands-trimet-set-the-standard-for-open-transit-data>, 5 janvier 2010, consulté le 19 octobre 2015.

ROUVROY Antoinette et BERNIS Thomas, « Gouvernamentalité algorithmique et perspectives d'émancipation. Le disparate comme condition d'individuation par la relation ? », *Réseaux*, 2013, vol. 1, n° 177, p. 163-196.

RUPPERT Evelyn, « The Governmental Topologies of Database Devices », *Theory, Culture & Society*, 2012, vol. 29, n° 4-5, p. 116-136.

SACK Robert David, *Human territoriality: its theory and history*, New York, Cambridge University Press, 1986.

SASSEN Saskia, *Territory, Authority, Rights: From Medieval to Global Assemblages*, Princeton, Princeton University Press, 2006, 502 p.

SAURUGGER Sabine et SUREL Yves, « L'eupéanisation comme processus de transfert de politique publique. », *Revue internationale de politique comparée*, 2006, vol. 13, n° 2, p. 179-211.

SCHWARTZ Olivier, « L'empirisme irréductible » dans *Le hobo : sociologie du sans-abri*, Paris, Nathan, 1993, p. 265-308.

SCOTT James, *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*, New Haven, Yale University Press, 1998, 445 p.

SEGUIN Ève, « Pourquoi les exoplanètes sont-elles politiques ? Pragmatisme et politicalité des

sciences dans l'œuvre de Bruno Latour », *Revue française de science politique*, 2015, vol. 65, n° 2, p. 279-302.

SHELLER Mimi et URRY John, « The New Mobilities Paradigm », *Environment and Planning A*, 2006, vol. 38, n° 2, p. 207-226.

SHELTON Taylor, ZOOK Matthew et WIIG Alan, « The “Actually Existing Smart City” », *Cambridge Journal of Regions, Economy and Society*, 2015, vol. 8, n° 1, p. 13-25.

SIBILLE Bastien, « Voir l'Europe pour la faire. Un système d'informations géographiques dans la “gouvernance” européenne. », *Politique européenne*, 2010, n° 31, p. 147-172.

SIMMEL Georg, *Les grandes villes et la vie de l'esprit*, Paris, Payot, 2013, 112 p.

SLEE Tom, *Seeing like a geek.*, <http://crookedtimber.org/2012/06/25/seeing-like-a-geek/> , 25 juin 2012, consulté le 6 octobre 2015.

SÖDERSTRÖM Ola, PAASCHE Till et KLAUSER Francisco, « Smart cities as corporate storytelling », *City*, 2014, vol. 18, n° 3, p. 307-320.

STAR Susan Leigh, « The Ethnography of Infrastructure », *American Behavioral Scientist*, 1999, vol. 43, n° 3, p. 377-391.

STAR Susan Leigh et GRIESEMER James R., « Institutional Ecology, “Translations” and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology », *Social Studies of Science*, 1989, vol. 19, n° 3, p. 387-420.

STAR Susan Leigh et RUHLER Karen, « Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces », *Information Systems Research*, 1996, vol. 7, n° 1, p. 111-134.

STAR Susan Leigh et STRAUSS Anselm, « Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work », *Computer Supported Cooperative Work*, 1999, vol. 8, n° 1-2, p. 9-30.

STRAUSS Anselm, *La trame de la négociation: Sociologie qualitative et interactionnisme*, Paris, L'Harmattan, 1992, 318 p.

STRAUSS Anselm, « The articulation of project work: an organizational process », *The Sociological Quarterly*, 1988, vol. 29, n° 2, p. 163-178.

STRAUSS Anselm, « Work and the division of labor », *The Sociological Quarterly*, 1985,

vol. 26, n° 1, p. 1-19.

THEVENOT Laurent, « Les investissements de forme » dans Laurent Thévenot (ed.), *Conventions économiques*, Paris, PUF, 1986, p. 21-71.

THEVENOT Laurent et DESROSIERES Alain, *Les catégories socio-professionnelles.*, 5e éd., Paris, La Découverte, 2002, 121 p.

TKACZ Nathaniel, « From open source to open government : a critique of open politics », *Ephemera: Theory and Politics in Organization*, 2012, vol. 12, n° 4, p. 386-405.

TOWNSEND Anthony, *Smart Cities - Big Data, Civic Hackers, and the Quest for a New Utopia*, New York, W. W. Norton & Company, 2013, 400 p.

TRANOS Emmanouil et GERTNER Drew, « Smart networked cities? », *Innovation: The European Journal of Social Science Research*, 2012, vol. 25, n° 2, p. 175-190.

TROMPETTE Pascal et VINCK Dominique, « Retour sur la notion d'objet-frontière », *Revue d'anthropologie des connaissances*, 2009, vol. 3, n° 1, p. 5-27.

TURNER Fred, *Aux sources de l'utopie numérique : De la contre-culture à la cyberculture, Stewart Brand, un homme d'influence*, Caen, C&F éditions, 2013, 430 p.

URRY John, *Mobilities*, Cambridge, Polity Press, 2007, 336 p.

URRY John, *Sociology beyond Societies. Mobilities for the twenty-First Century*, London, Routledge, 2000, 268 p.

VANOLO Alberto, « Smartmentality: The Smart City as Disciplinary Strategy », *Urban studies*, 2014, vol. 51, n° 5, p. 883-898.

VICKERY Graham, « Review of recent studies on PSI re-use and related market developments. », *Information Economics*, 2011, p. 41.

VOUTAT Bernard, « A propos de l'objet de la science politique : sens et non-sens d'une question récurrente » dans Pierre Favre, Olivier Fillieule et Fabien Jobard (eds.), *L'atelier du politiste. Théories, actions, représentations.*, Paris, La Découverte, 2007, p. 35-57.

WEBER Max, *Economie et société*, Paris, Pocket (coll. « Agora »), 1995.

WEINBERGER David, *Too Big to Know*, New York, Basic Books, 2011, 256 p.

WIIG Alan, « IBM's smart city as techno-utopian policy mobility », *City*, 2015, vol. 19, n° 2-3, p. 258-273.

Testing governance: the laboratory lives and methods of policy innovation labs., [Working paper], 2015.

WORTHY Ben, « The Impact of Open Data in the UK: Complex, Unpredictable and Political », *Public Administration*, 2015, vol. 93, n° 3, p. 788-805.

YU Harlan et ROBINSON David, « The New Ambiguity of "Open Government" », *UCLA Law Review Discourse*, 2012, vol. 59, n° 178, p. 180-208.

ZIMMERMAN Ann, « Not by metadata alone: the use of diverse forms of knowledge to locate data for reuse », *International Journal on Digital Libraries*, 2007, vol. 7, n° 1, p. 5-16.

ZITOUN Philippe, *La fabrique politique des politiques publiques. Une approche pragmatique de l'action publique.*, Paris, Presses de Sciences Po, 2013, 340 p.

Ambition numérique. Pour une politique française et européenne de la transition numérique. Rapport remis au Premier Ministre., Paris, Conseil national du numérique, 2015.

Ouverture des données de transport, Paris, Rapport remis au secrétaire d'État chargé des Transports, de la Mer et de la Pêche, 2015.

Pourquoi n'y a-t-il pas de consensus sur une licence Open Data en France?, <https://libertic.wordpress.com/2011/07/05/pourquoi-ny-a-t-il-pas-de-consensus-sur-une-licence-open-data-en-france/>, 5 juillet 2011, consulté le 10 août 2010.

Licences OpenData: L'APIE grille la priorité à ÉtaLab et invente le pseudo-libre, <http://www.regardscitoyens.org/licences-opendata-lapie-grille-la-priorite-a-etalab-et-invente-le-pseudo-libre/>, 19 juillet 2010, consulté le 10 août 2010.

L'OpenData, c'est bien plus que de la communication politique!, <http://www.regardscitoyens.org/lopendata-cest-bien-plus-que-de-la-communication-politique/>, 9 juin 2010, consulté le 10 août 2015.

V. Liste des entretiens

70 entretiens semi-directif ont été réalisés dans le cadre de cette thèse. Les entretiens sont classés par pays d'enquête puis par institution et enfin par ordre chronologique. Certaines personnes ont été rencontrées à plusieurs reprises.

A. *France*

1) **Grand Lyon**

1. Chef de projet informatique, en charge de l'open data, 22 novembre 2012 & 11 février 2015
2. Chef de projet utilisateur, en charge de l'open data, 23 novembre 2012 & 17 février 2015
3. Chef de projet services numériques, 6 décembre 2012
4. Responsable du service Information géographique, 11 décembre 2012 & 15 janvier 2014
5. Chargé de mission à la direction de la prospective, 17 décembre 2012
6. Chargé de mission mobilité urbaine, 8 janvier 2013 & 19 janvier 2015
7. Directeur du projet ville intelligente, 14 janvier 2013
8. Directeur des systèmes d'information, 16 janvier 2013
9. Architecte des système d'information, 9 juillet 2013
10. Responsable de la plateforme open data, ex-administratrice des données de la direction de la propreté, 13 octobre 2013
11. Responsable données et maquettes 3D, 21 novembre 2013
12. Administrateur de données de la direction de la propreté, 22 novembre 2013
13. Responsable de l'unité Données de référence, 26 novembre 2013 & 18 février 2015
14. Administrateur de données à la direction de l'eau, 4 décembre 2013
15. Administrateur de données à la direction du développement économique, 5 décembre 2013
16. Administrateur de données à la direction des politiques d'agglomération, 6 décembre 2013
17. Responsable du service observation et valorisation des données, 14 janvier 2014

18. Administrateur de données à la direction de la voirie, 17 janvier 2014
19. Responsable du service mobilité urbaine, 22 janvier 2014
20. Responsable du service MOI, direction de l'aménagement, 20 février 2014
21. Responsable du pôle technique du PC Criter, 9 mars 2014
22. Responsable du pôle Infotrafic, 12 mars 2014
23. Chef de projet, service déplacements, 11 janvier 2015
24. Chargée de mission au sein de la direction des politiques d'agglomération, 19 janvier 2015
25. Responsable administration des données, 24 février 2015
26. Directeur des systèmes d'information, 18 mars 2015
27. Responsable de l'unité Exploitation / Information du PC Criter, 20 mars 2015
28. Administratrice générale des données, 21 octobre 2015
29. Vice-présidente Innovation, Métropole intelligente, développement numérique, 22 octobre 2015

2) Producteurs et/ou utilisateurs de données

1. Responsable Open data, Région PACA, 15 mai 2013
2. PhD Candidate, IBM, 24 février 2014 (entretien téléphonique)
3. Administrateur général des données, SGMAP, 20 mars 2014
4. Ingénieur, Géovélo, 7 juillet 2014 (entretien téléphonique)
5. Chef de projet, Cityway, 9 juillet 2014 (entretien téléphonique)
6. Responsable du SIG au sein de la mairie de Vaulx-en-Velin, 24 juillet 2014
7. Directeur des systèmes d'information adjoint, Aéroports de Lyon, 25 juillet 2014 (entretien téléphonique)
8. Responsable du SIG à la ville de Lyon, 25 juillet 2014
9. Responsable du SIG au sein de la mairie de Tassin-la-demi-lune, 26 juillet 2014
10. Responsable du SIG au sein de la mairie de Chassieu, 28 juillet 2014
11. Responsable du SIG au sein de la mairie de Caluire et Cuire, 29 juillet 2014
12. Responsable du SIG au sein de la mairie de Vénissieux, 31 juillet 2014
13. Responsable du SIG au sein de la municipalité de Villeurbanne, 7 août 2014
14. Chef de projet informatique, Keolis Lyon, 13 août 2014

15. Directeur des applications grand public, JC Decaux, 27 août 2014 (entretien téléphonique)
16. Analyste géographe, Nokia Here, 9 janvier 2015
17. Manager, Oslandia, 26 janvier 2015
18. Chef de projet, IBM, 27 janvier 2015
19. Responsable marketing et communication, Canal TP, 28 janvier 2015
20. Chef de produit, Google, 9 février 2015 (entretien téléphonique)
21. Senior Policy Manager, Google, 11 février 2015 (entretien téléphonique)
22. Chef de projet, Renault Trucks, 13 février 2015 (entretien téléphonique)
23. Responsable SIG, Agence d'urbanisme de Lyon, 25 février 2015
24. Président d'OpenStreetMap France, Coordinateur de la Base adresse nationale, Etalab, 4 mars 2015
25. Data analyst, Canal TP, 4 mars 2015
26. Ingénieur SIG, Forcity, 5 mars 2015

B. Etats-Unis

1) New York

1. Director of Public Affairs, Mayor's Office of Data Analytics, New York City, 30 octobre 2014
2. Director & co-Founder, Data-Pop Alliance, 6 novembre 2014
3. Co-Founder and Chief Operating Officer, Placemeter, 10 novembre 2014
4. Chief Technology Officer, The Governance Lab at NYU, 11 novembre 2014
5. Director of OpenNY, Office of the Governor, New York State, 24 novembre 2014
6. City Planner/GIS Specialist at NYC Department of City Planning, 24 novembre 2014
7. Project Manager, Health Data NY, 10 décembre 2014
8. Civic Engagement Manager at NYC DoITT, 15 décembre 2014

2) Boston

1. Deputy Director of Operations Technology at Massachusetts Bay Transportation Authority, 11 décembre 2014
Special Assistant for Strategic Initiatives at Massachusetts Bay Transportation Authority, 11 décembre 2014
2. Co-Founder, Mayor's Office of New Urban Mechanics, City of Boston, 11 décembre 2014
Principal Data Scientist, City of Boston, 11 décembre 2014

3) Philadelphie

1. Chief Data Officer at City of Philadelphia, 4 décembre 2014
2. Brigade Captain, Code for Philly, 4 décembre 2014
3. Director, Mayor's Office of New Urban Mechanics, City of Philadelphia, 5 décembre 2014
4. Co-Director, Mayor's Office of New Urban Mechanics, City of Philadelphia, 5 décembre 2014
5. Senior Project Leader at Southeastern Pennsylvania Transportation Authority (SEPTA), Philadelphie, 5 décembre 2014

