

Antoine COURMONT

Résumé

**Politiques des données urbaines.
*Ce que l'open data fait au gouvernement urbain***

Thèse de science politique

Sous la direction de Dominique Boullier

Soutenue le 16 décembre 2016

Jury :

M. Dominique BOULLIER, Professeur des Universités en sociologie, 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 (président)

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

Mots-clefs : open data, donnée urbaine, sociologie des données, numérique, smart city, gouvernement urbain, Lyon

Si les administrations municipales ont toujours été productrices et utilisatrices de savoirs de gouvernement, d'indicateurs ou de statistiques, à partir du milieu des années 2000, de nouvelles entités ont fait leur apparition dans le gouvernement urbain : les données. Elles sont maintenant devenues un enjeu central pour de nombreuses métropoles, qui développent des stratégies, mettent en œuvre des instruments, transforment leur organisation et créent de nouveaux métiers pour gérer cette entité informationnelle. Comment les données sont-elles devenues un enjeu de gouvernement ? Qu'est-ce que l'émergence de ces nouvelles entités fait au gouvernement urbain ?

Quand les données arrivent en ville : le libéralisme informationnel à l'échelle urbaine

L'émergence des données dans le gouvernement des villes est étroitement liée au développement des démarches d'*open data* et de *smart cities* au sein de nombreuses métropoles. La mise en circulation des données, le passage d'une logique de stock à une logique de flux, met en lumière l'application des principes du libéralisme informationnel à l'échelle urbaine.

L'*open data* est un mouvement international, initié dans les pays anglo-saxons, visant à mettre en circulation les données des administrations publiques afin de les rendre accessibles aux citoyens. La mise à disposition des données doit favoriser à la fois le développement économique et la transparence et la participation citoyenne. En France, la ville de Rennes est la première collectivité à s'engager dans une politique d'*open data* dès 2010. Aujourd'hui, plus d'une centaine de collectivités sont engagées dans des démarches d'ouverture de leurs données. Ce chiffre devrait s'étendre dans les prochains mois, puisque la loi NOTRE, votée en 2015, impose l'ouverture des données à l'ensemble des collectivités de plus de 3500 habitants.

Parallèlement, sous l'impulsion d'entreprises des technologies de l'information telles que IBM¹, la *smart city* a fait son apparition dans les métropoles. Si la ville intelligente reste une notion floue, on peut la définir de manière générique comme l'usage généralisé de données numériques pour rationaliser la planification et la gestion des villes². 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 villes³. Pour cela, il est nécessaire de décroiser les « silos » de données et passer d'une logique de stock à une logique de flux, c'est-à-dire de mettre en circulation les données.

¹ Ola Söderström, Till Paasche et Francisco Klauser, « Smart cities as corporate storytelling », *City*, 2014, vol. 18, n° 3, p. 307-320.

² Anthony Townsend, *Smart Cities - Big Data, Civic Hackers, and the Quest for a New Utopia*, New York, W. W. Norton & Company, 2013, 400 p.

³ 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.

Open data et *Smart city* traduisent l'application de l'idéologie du « libéralisme informationnel » à l'échelle des villes. Théorisé par Benjamin Loveluck à partir d'une généalogie politique d'Internet, le libéralisme informationnel est une reformulation de la théorie politique libérale fondée sur les flux d'informations au sein des réseaux informatiques⁴. L'information doit être libre afin que ces flux d'informations résolvent de multiples problèmes dans une logique cybernétique. Le paradigme de la mobilité, mis en avant par John Urry⁵, 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.

Suivre la chaîne de la donnée : approche théorique et méthodologique

Au travers de l'analyse de la politique d'*open data* de la Métropole de Lyon, cette thèse a pour ambition de suivre au plus près les données pour analyser ce que leur mise en circulation fait au gouvernement urbain. 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. L'approche théorique poursuivie mobilise la sociologie urbaine, l'analyse de l'action publique et la sociologie des sciences et des techniques (STS), en particulier la sociologie de la quantification développée en France par Alain Desrosières⁶. En croisant les analyses internalistes et externalistes des données, la thèse étudie les recompositions des modalités de gouvernement des espaces urbains.

Cette sociologie des données 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 : 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. 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 gouverner. A chacune de ces étapes, la thèse pointe les politiques des données urbaines, entendues comme les modes d'association et de composition de collectifs avec et autour des données. Suivre la chaîne de la donnée permet de mettre en évidence 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.

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. Cette observation ethnographique a été effectuée au sein de la

⁴ Benjamin Loveluck, *Réseaux, libertés et contrôle. Une généalogie politique d'internet*, Paris, Armand Colin, 2015.

⁵ John Urry, *Mobilities*, Cambridge, Polity Press, 2007, 336 p.

⁶ Alain Desrosières, *La politique des grands nombres. Histoire de la raison statistique.*, Paris, La Découverte, 1993.

Métropole de Lyon dans le cadre d'un dispositif CIFRE. De novembre 2012 à juin 2015, j'ai pu suivre la politique d'ouverture des données *en train de se faire* en accédant aux coulisses de l'administration et aux arènes discrètes difficilement perceptibles par des sources externes. 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 (Boston, New York, Philadelphie).

Argumentaire de la thèse

La thèse est organisée en cinq chapitres qui correspondent à cinq étapes de la chaîne de la donnée ouverte.

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 les données sont constituées de multiples attachements qui les rendent indissociables des vastes réseaux sociotechniques dans lesquels elles sont insérées. Attachées, les données sont également productrices d'attachements. Ces liens les dotent d'un pouvoir d'agir, d'une *agency* : en mettant en relation d'une façon particulière les utilisateurs et l'espace urbain, elles participent à l'émergence d'agencements spécifiques associant l'institution communautaire 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 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 s'intéresse aux modalités de réutilisation des données ouvertes. Pour dépasser leur cadrage originel, il est nécessaire de traduire ces données sans nuire à leur solidité.

Par la description des modalités d'attachement des données à de nouveaux utilisateurs, le pluralisme des politiques de réutilisation est mis en avant. Ces politiques 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 cinquième 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.

Principaux résultats et apports opérationnels

Qu'est-ce qu'une donnée ?

Le premier apport de cette thèse est de proposer une définition sociologique de la donnée comme un dispositif sociotechnique, constitué d'un réseau composé d'un ensemble d'actants reliés entre eux par des associations particulières. Cette définition insiste ainsi sur le caractère relationnel de la donnée : la donnée est constituée d'un tissu de relations. Analyser l'évolution de ces relations, c'est continuellement redéfinir ce qu'est ou n'est pas une donnée. La donnée n'a ainsi pas de caractéristiques essentielles. 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.

Qu'est-ce que l'open data ?

Le second 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.

De manière plus générale, j'affirme dans la thèse que l'open data n'est qu'une des manifestations d'un mouvement plus large de mise en œuvre, à l'échelle urbaine, des principes du libéralisme informationnel, tel que les a théorisés Benjamin Loveluck. Ce libéralisme informationnel se manifeste par l'injonction à une mise en circulation généralisée des données, illustrée par l'open data, mais plus généralement par le développement des *smart cities*, qui visent à dépasser les silos de données pour construire une nouvelle couche informationnelle. Ce passage d'une logique de stock à une logique de flux de données, étudié empiriquement dans cette thèse au travers du cas de l'open data, interroge sur les modalités de recomposition du gouvernement de la ville à l'ère du numérique.

Qu'est-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 en mettant en évidence deux processus liés à la mise en circulation des données urbaines : d'une part, la donnée devient un enjeu de gouvernement ; d'autre part, les frontières sectorielles, institutionnelles et territoriales sont questionnées.

Durant les trois ans de mon étude, de 2012 à 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.

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 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 plus généralement 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 ces flux, il perd la maîtrise de ses politiques publiques. L'exemple des GPS automobiles cité dans le chapitre 4 est révélateur de ce risque. Dans l'incapacité de réguler les flux de données,

la collectivité perd la maîtrise de sa politique de régulation du trafic routier. Indépendamment des pouvoirs publics, des acteurs peuvent réguler la ville par le biais de flux informationnels.

D'autre part, la mise en circulation des données questionne les frontières sectorielles, institutionnelles et territoriales. 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.