

Le tournant numérique des données immobilières. Permanences et recompositions.

Résumé

L'immobilier se conjugue au pluriel. Il n'y a pas un mais plusieurs marchés du logement. La segmentation reproduit la chaîne de production de l'offre, entre construction et commercialisation, achat et location. Elle renvoie à l'organisation institutionnelle et professionnelle du secteur, entre public et privé, neuf et ancien, promotion et intermédiation immobilières, logement individuel et collectif. Elle a une dimension spatiale, fonction de la diversité des territoires, à l'échelle nationale comme au niveau local.

Les données immobilières reproduisent cette mosaïque. Elles résultent d'un processus de production qui s'inscrit dans la diversité des cadres d'action des acteurs de l'immobilier. Elles ne sont jamais « données » mais toujours construites. Cette étude ne les prend pas toutes pour objet mais s'arrête aux données de prix et de loyer des logements du marché privé. Celles-là sont les plus stratégiques. Elles font l'intérêt de tous : Etat, élus, collectivités, particuliers, promoteurs, agents immobiliers, investisseurs, financeurs de l'offre et de la demande, médias, opérateurs de données, start-ups immobilières, etc.. Et répondent à de nombreux enjeux : d'information et de transparence du marché ; d'accompagnement du marché et d'évaluation des politiques publiques ; de veille concurrentielle et de décision stratégique ; de suivi de la construction, secteur d'activité déterminant au plan macro-économique ; de financement de l'offre et de solvabilisation de la demande ; d'expertise et de gestion d'actif ; d'innovation en termes de service, etc..

De fait, les données de prix et de loyer sont des données de marché. Elles renvoient à des flux de logements, pas à des stocks ; elles visent des biens échangés, pas des biens occupés. Elles indiquent le point de rencontre de l'offre et de la demande au moment de l'échange. Autre particularité, les biens immobiliers se présentent sur le marché en faible proportion relativement à l'ensemble du parc et à intervalle de plusieurs années, voire décennies. Leur valeur est étroitement liée au contexte conjoncturel de leur commercialisation mais elle est connue, après enregistrement, dans des délais qui se comptent en semaines ou en mois.

Les données de prix et de loyer apparaissent d'abord éclatées, entretenant le flou qui entoure les dynamiques du marché privé du logement. Mais comme le segment auquel elles se rapportent, elles forment en réalité un ensemble circonscrit et structuré. C'est un écosystème fait de complémentarités, de concurrences et d'interdépendances tout à la fois. Cette étude tend à le démontrer en rentrant dans la fabrique de la donnée aujourd'hui. Le travail a consisté à effectuer le relevé des données existantes et à reconstituer leur chaîne de production, de la collecte à la diffusion en tenant compte du cadre d'action, déterminant, de leurs producteurs. Il est parti des données de première main, résultats d'une démarche qui, remontant à la source, a vocation à renseigner la valeur des biens échangés sur le marché privé. L'ensemble est riche mais pas dépourvu de lacunes, laissant plusieurs segments dans l'angle mort de l'observation immobilière.

Les recherches ont suivi le fil de la question numérique. La révolution numérique a déjà saisi la ville et elle rattrape l'immobilier et la construction. Mais percute-t-elle la production même des données de prix et de loyer ? Introduit-elle de nouvelles sources, de nouveaux outils et de nouveaux acteurs dans l'écosystème ?

Le premier tour d'horizon effectué a été celui d'un écosystème de taille relativement petite et plutôt stable dans ses composantes et ses méthodes. Ainsi :

- le monde de la production des données de prix et de loyer en France compte une dizaine d'acteurs, pas plus ;
- parmi les producteurs de données, l'immense majorité est en place depuis plusieurs décennies, laissant la Proptech (des start-ups ambitionnant de réinventer les services immobiliers) à une position encore marginale, sinon minoritaire ;
- la puissance publique joue un rôle central dans l'écosystème, non seulement en fournissant la donnée légitime mais en conditionnant la production générale à son ouverture ;
- le numérique s'impose à tous les maillons de la chaîne mais laisse encore inchangées de nombreuses pratiques de production et de diffusion.

Ces constats allant contre l'hypothèse de disruption vérifiée dans d'autres secteurs d'activité tels que la mobilité (et ses données), il a fallu les interroger et les mettre à l'épreuve d'un point de vue inverse sur l'objet de l'étude : celui des utilisateurs de données plutôt que celui des producteurs (à l'exclusion des producteurs qui avaient déjà été entendus en tant qu'utilisateurs bien sûr). En particulier, un complément d'enquête a été effectué auprès des professionnels de l'immobilier et a consisté à identifier le type de données qu'ils consommaient.

Les entretiens effectués ont permis de confirmer le rôle central des producteurs historiques de prix et de loyers dans l'information aux professionnels. Ils ont également révélé le recours à des solutions nouvelles. Ces solutions sont proposées par des « agrégateurs », nouveaux entrants qui empruntent à la *data visualisation* et/ou au *machine learning* (apprentissage automatique) pour développer *dataplaces* et/ou estimations automatiques. Tandis que les premiers concentrent des données éclatées et offrent des outils simples et agiles pour les consulter toutes à la fois, les seconds répondent à la demande de données sur mesure et stratégiques, voire prédictives. Tous surfent sur la vague servicielle du numérique et laissent une part majeure à l'utilisateur, au moyen d'outils reposant sur des fonctions essentiellement automatiques.

A ce stade, les agrégateurs sont mobilisés en complément des producteurs historiques de données de prix et de loyer et à titre exploratoire. Ce constat est nouveau et la rapidité des transformations en cours pourrait impliquer de le réviser prochainement. Plus généralement, les résultats et analyses présentés dans ce rapport ont dû être actualisés plusieurs fois jusqu'au point final, tant les évolutions observées au fil de l'eau ont été rapides et nombreuses. Ils saisissent la situation en avril 2019, quelques jours seulement après l'ouverture à tous des données quasi exhaustives sur les transactions immobilières détenues par la Direction générale des finances publiques (DGFIP) : les données DVF (Demandes de valeurs foncières), du nom du service qui assure leur accès.

L'événement est de taille. Il promet de combler les lacunes des données existantes et leur indisponibilité sous des formats et à des conditions d'accès susceptibles de répondre à la demande. Non seulement il réaffirme le rôle central de la puissance publique au sein de l'écosystème, mais il touche directement les particuliers, en se soldant par la mise en service par Etalab d'une carte interactive de consultation des transactions immobilières. L'ensemble révèle la

puissance potentiellement disruptive du public. A l'heure de la PropTech, qui est-ce qui percute le système : les start-ups immobilières ou l'Etat ? Les agrégateurs ou les producteurs historiques de données ? Si tout porte à croire que les constats réunis dans ce rapport seront vite dépassés, l'ambition est ici, pour répondre à ces questions, de fournir le cadre nécessaire à une meilleure compréhension des évolutions à venir et des enjeux qui les sous-tendent.

Le présent rapport prend place dans un projet comparatif. Il est complété d'un panorama sur les principales données existantes dans huit pays : en Allemagne, Angleterre, Espagne, Italie, Norvège, Suède, Suisse et aux Etats-Unis. Ce panorama ne se veut pas exhaustif mais indicatif, à titre de comparaison avec la France, de différents écosystèmes de production de données. Il s'organise autour des trois questions suivantes : 1) la puissance publique est-elle, comme en France, un acteur majeur de la production des données immobilières ? ; 2) quels pays ouvrent-ils leurs données immobilières ? ; 3) quelle place les acteurs privés occupent-ils dans l'écosystème de production ?