



HAL
open science

”Transparence des marchés immobiliers : apports et limites de l’Open Data”, Interview dans Politiquedulogement.com

Alexandre Coulondre, Vincent Lasserre-Bigorry

► **To cite this version:**

Alexandre Coulondre, Vincent Lasserre-Bigorry. ”Transparence des marchés immobiliers : apports et limites de l’Open Data”, Interview dans Politiquedulogement.com. 2023. hal-04937783

HAL Id: hal-04937783

<https://hal.science/hal-04937783v1>

Submitted on 10 Feb 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Transparence des marchés immobiliers : apports et limites de l'Open Data

Alexandre Coulondre^[1] et **Vincent Lasserre-Bigorry**^[2] ont publié en octobre 2022 pour l'institut IDHEAL une étude intitulée « [Les territoires de la \(non-\)production de logements](#) » dans laquelle ils élaborent une géographie de la production de logements à l'échelle des intercommunalités du territoire au cours des années 2010 à 2019.

Cette étude mobilise dix sources de données relativement nouvelles, ouvertes à tous et fournies au minimum à l'échelle communale, voire sous forme exhaustive. C'est l'occasion pour [poiltiquedulogement.com](#) de les interroger sur le processus d'ouverture au public de ces ressources, de leurs apports et de leurs limites.

Alexandre Coulondre est également co-auteur avec Claire Juillard de deux rapports sur le sujet :

- Coulondre A. et Juillard C. (2022) [Le marché de la donnée au service de la transparence du marché du logement ?](#) Working Paper, Science-Po Chaire Villes et numérique, n°01-2022.
- Coulondre A. et Juillard C. (2021) [Ce que la data fait au marché](#). PUCA, LIFTI, BPCE, Crédit Logement.

Vos travaux sur l'immobilier se caractérisent par l'utilisation de sources de données mises à la disposition du public depuis quelques années sous forme de bases exhaustives : DVF, SITADEL, RPLS, MAJIC personnes morales, BDNB ...

Alexandre Coulondre : En effet, beaucoup de nos travaux de recherche n'auraient pas été possibles il y a quelques années, et notamment celui que nous venons de réaliser pour IDHEAL sur les territoires de la production et de la non-production de logements. L'objectif de cette recherche était de caractériser finement la production résidentielle dans une approche multivariée et à une échelle locale. Et il s'agissait aussi de prendre en compte les contextes territoriaux dans lesquels se réalise cette production en ce qui concerne le peuplement, les finances locales ou encore l'artificialisation des sols.

Pour réaliser une telle recherche, nous avons dû mobiliser et croiser dix sources de données ouvertes. SITADEL2 (recensement des permis de construire) a été au cœur de l'analyse. Cette source est accessible depuis longtemps sous forme d'indicateurs communaux mais l'accès à sa version détaillée est récent. Nous avons aussi mobilisé : demande de valeurs foncières (DVF pour les transactions), le répertoire du parc locatif social (RPLS), le recensement de la population, l'observatoire national de consommation des espaces naturels et forestiers (OENAS), le recensement des éléments d'imposition (REI), etc. Sans les vagues récentes d'ouverture des données cela n'aurait pas été possible.

En cela, l'Open Data constitue une avancée majeure pour la recherche sur les marchés immobiliers et les dynamiques territoriales. Les chercheurs n'avaient pas été intégrés historiquement dans la liste des ayants droit de la donnée publique^[3]. Ils ont dû attendre l'Open Data de la fin des années 2010.

Vincent Lasserre-Bigorry : Il faut préciser aussi que l'ouverture des sources que nous venons de citer est fondamentalement liée à l'ouverture des référentiels géographiques associés. Par exemple, DVF est structurée sur le référentiel cadastral, quand les données du recensement sont conçues à la maille des IRIS (Îlots Regroupés pour l'Information Statistique) et le RPLS à l'adresse. L'articulation des

sources nécessite l'articulation des référentiels géographiques. L'ouverture concomitante du PCI (Plan Cadastral Informatisé), de la BAN (Base Adresse Nationale), ou encore il y a quelques années des fonds de cartes des IRIS est ce qui a rendu réellement possible l'exploitation des données urbaines et immobilières par ailleurs libérées.

Quelle est l'histoire récente de cette mise à disposition de bases de données immobilières ?

Alexandre Coulondre : L'ouverture de données brutes est une pratique qui concerne aujourd'hui tous types d'acteurs. Les collectivités territoriales, le notariat, ou encore les entreprises ouvrent des données. Mais l'Etat joue un rôle central dans ce processus.

Il faut d'ailleurs remarquer que la question des données immobilières, qui nous intéresse ici, n'est qu'un volet d'un mouvement plus général d'ouverture des données publiques qui a pris de l'ampleur depuis environ une dizaine d'années en France. La loi *Pour une République Numérique* de 2016 a fixé un cadre qui facilite les ouvertures dans de nombreux domaines. Un cas emblématique est celui des données de transport (horaires des trains, etc.) qui ont été parmi les premières à faire l'objet d'une ouverture massive et d'une utilisation par des acteurs privés.

Sur la question de l'immobilier, l'acte fondateur a été l'ouverture des données DVF en 2019 à la faveur de la loi ESSOC^[4] votée un an plus tôt. Comme nous l'avons dit, ces données étaient déjà accessibles depuis 2010 mais seulement pour certains acteurs publics. Le secret fiscal justifiait cette limitation. Pourtant en 2019, et alors que les normes européennes sur la gestion des données personnelles se durcissaient avec le règlement général sur la protection des données (RGPD), il y a eu une très forte volonté gouvernementale de diffuser ces données fiscales dans un format brut et en conservant leur précision géographique. De fait, la donnée est fournie aujourd'hui à la transaction avec reprise du numéro de parcelle. Ça a été un vrai tournant.

Vincent Lasserre-Bigorry : Et très vite après DVF d'autres sources de données ont été ouvertes comme les données du ministère en charge du logement que sont SITADEL2 sur les permis de construire, ou RPLS sur le parc de logements sociaux. Plus récemment, on peut évoquer celles de l'ADEME sur les DPE (Diagnostics de Performance Energétique) ou celle du CSTB sur les bâtiments et leurs consommations (BDNB).

Qu'est-ce qui justifie, pour l'Etat, la mise à disposition de données immobilières et urbaines à tous types de publics ?

Vincent Lasserre-Bigorry : Il y a déjà une contrainte réglementaire. En termes d'expropriation par exemple, l'Union Européenne a consacré un principe d'égalité des armes entre les parties qui implique que tous les acteurs, qu'ils soient privés ou publics, puissent avoir un même niveau d'information.

Alexandre Coulondre : il y a aussi une adhésion implicite à la pensée économique que l'on pourrait qualifier d'orthodoxe qui voit les territoires comme des marchés pouvant être efficaces à la condition qu'ils soient « transparents ». Théoriquement cette transparence devrait être garantie par l'existence de données détaillées et facilement accessibles. C'est un raisonnement théorique qui fait beaucoup débat dans le monde académique mais qui influence dans les faits les politiques d'ouverture des données immobilières. Une des ambitions de ces politiques est de peser indirectement sur les prix immobiliers. En théorie, la bonne information de tous doit limiter une partie des effets inflationnistes causés par l'asymétrie d'information.

Si on se concentre sur les enjeux de l'étude et de la recherche, quelles sont les exploitations nouvelles qui sont désormais permises par ces ouvertures ?

Alexandre Coulondre : Plus les données sont disponibles et brutes, plus les potentiels analytiques sont nombreux. On est loin d'avoir encore exploité tout ce potentiel, d'autant que de nouvelles sources arrivent encore chaque trimestre.

Je dirais néanmoins que l'Open Data renouvelle l'approche au moins sur deux aspects. Le premier est lié au géoréférencement très précis associé aux données brutes. Cela permet de descendre à des échelles fines comme la commune, le quartier, l'adresse ou la parcelle. Ces échelles permettent de saisir toute la complexité de la fragmentation des territoires.

Le deuxième aspect renvoie à l'ouverture des référentiels géographiques que nous avons déjà évoqué et qui permet l'interopérabilité des sources. Il est techniquement envisageable aujourd'hui d'articuler des sources de données que l'on étudiait auparavant de façon cloisonnée. Par exemple, on peut reconstituer l'histoire des programmes immobiliers en retrouvant dans DVF (données sur les transactions) les prix de vente associés aux terrains et aux logements situés à l'intérieur des périmètres de projets identifiés dans SITADEL2 (données sur les permis de construire). On passe ainsi d'une étude de la fixation des prix sur chaque marché à une étude de la chaîne de valeur globale en immobilier.

En cela, la diffusion des données ne change pas seulement la précision des mesures mais peut changer aussi les débats conceptuels et, par extension, les débats politiques sur les régulations qui s'y rapportent.

Vincent Lasserre-Bigorry : De même, le géoréférencement très précis des données permet d'enrichir les sources classiques de la statistique publique, comme le recensement, et décrire de manière bien plus fine les territoires dans leurs différentes dimensions.

Ces nouvelles bases de données exhaustives changent-elles la donne aussi pour les études et observatoires locaux ?

Alexandre Coulondre : Les acteurs locaux de l'observation comme les collectivités territoriales ne sont pas fondamentalement affectés par ce mouvement d'ouverture dans la mesure où ils faisaient partie de la liste des ayants droit historiques de la donnée publique et avaient déjà accès à des sources comme DVF depuis le début des années 2010 *via* des procédures spéciales d'obtention.

L'Open Data a néanmoins constitué un changement. Parfois pour le meilleur car la donnée ouverte est disponible sur tous les territoires alors qu'auparavant les ayants droit avaient accès à la donnée uniquement sur leur territoire de compétences. Dès lors, les comparaisons interterritoriales sont facilitées.

Et aussi parfois pour le pire, parce l'Open Data s'accompagne souvent d'une baisse de la qualité de l'information. Pour des questions de respect des données personnelles, les sources ouvertes sont généralement tronquées. Des variables disparaissent au moment de l'ouverture comme ce fut le cas avec DVF où les identifiants des locaux et le régime fiscal associé à une transaction ne sont plus renseignés, ce qui pose des problèmes d'interprétation. Il est par exemple impossible de savoir désormais si la TVA est incluse dans le montant d'une mutation. Le passage à l'Open Data implique donc une certaine baisse de qualité de l'information pour les acteurs publics locaux de l'observation.

Ces derniers se sont alors reportés sur la base DV3F proposée par le CEREMA qui enrichit l'information de DVF grâce aux Fichiers Fonciers, une source qui est encore réservée aux ayants droits publics. Ceci a recréé en quelque sorte un modèle à deux vitesses dans l'accès aux données.

Justement, quelles sont les limites de ces données ouvertes et réputées exhaustives ?

Alexandre Coulondre : Quand on fait de la statistique on sait que toute donnée a des limites. L'enjeu est surtout de comprendre ces limites pour réaliser des traitements adaptés.

En l'occurrence, on peut regretter comme nous l'avons dit que la donnée soit souvent tronquée par rapport à sa version historique. On peut aussi regretter parfois un certain manque d'exhaustivité, notamment avec SITADEL2 dont le système de collecte ne parvient pas à ce jour à recenser la totalité des permis de construire.

Pour certaines utilisations, notamment professionnelles, on peut regretter enfin les délais assez longs de mise à disposition. Les analyses sur périodes récentes sont limitées. Dans la plupart des cas, l'utilisateur devra se contenter de travailler avec un recul d'un an, voire deux. Cette limite s'est fait particulièrement ressentir il y a deux ans quand il s'est agi de commencer à étudier l'impact de la crise du Covid-19 sur le secteur résidentiel. Dans un travail que j'ai réalisé avec Claire Juillard et Marianne Bléhaut, nous avons dû délaissé l'Open Data et recourir à des données de plateformes numériques. En l'occurrence, nous avons établi un partenariat avec le site *leboncoin* afin d'exploiter ses données de navigation^[5]. Celles-ci ont l'avantage d'être disponibles quasiment en temps réel.

Vincent Lasserre-Bigorry : Il y a aussi le fait que l'organisation des données et leur format dépend très fortement de l'utilisation qui en est faite par l'administration d'origine. DVF est produite par la DGFIP (direction générale des Finances Publiques). Cette origine fiscale est perceptible à quiconque a déjà ouvert le fichier. Les biens sont décomposés en plusieurs sous-entités pertinentes pour l'administration fiscale mais assez peu pour l'analyse économique des marchés. Un lourd travail de mise en forme et de « rétro-ingénierie » est nécessaire pour reconstituer un fichier exploitable.

Il faut ajouter que les conditions de production de certains fichiers ne sont pas toujours connues, les métadonnées sont souvent pauvres et certaines variables peuvent donc être difficilement exploitables notamment par méconnaissance de leurs spécificités et limites.

C'est à mon sens le cas de la base ouverte SIRENE, qui est une source administrative très riche permettant de connaître l'évolution des activités économiques d'un territoire à travers la démographie de ses établissements économiques. La fiabilité de cette base est fortement liée aux délais disparates de mise à jour des informations sur les établissements. C'est là que des erreurs peuvent être faites si les utilisateurs en font un usage trop rapide.

Alexandre Coulondre : On peut dire finalement que l'Open Data a baissé le coût d'accès aux données mais d'un autre côté a nettement relevé les coûts liés à la formation et à l'acquisition du savoir-faire.

Comment comprendre que les données soient diffusées sous une forme « brute », très peu, voire pas, retraitées ?

Alexandre Coulondre : C'est en effet assez paradoxal à première vue de rechercher la transparence des marchés et de fournir des données brutes et difficiles à utiliser.

Une explication renvoie aux moyens limités dont disposent les administrations qui ouvrent des données. Bien souvent, la mise à disposition de données n'est qu'une mission annexe dans la feuille de route de ces administrations. Mettre en forme et mettre en ligne des données représente un certain coût. Dès lors, elles ne sont pas incitées à aller plus loin, que ce soit dans la standardisation de

leurs données, ou dans la production d'expertise à destination du grand public (très peu d'administrations qui diffusent des données proposent aussi des outils de connaissance ou de visualisations par ailleurs).

Vincent Lasserre-Bigorry : C'est l'exemple que nous prenions avec la DGFIP, le fichier DVF tel qu'il est produit n'est vraisemblablement pas pensé pour être facilement exploitable mais plutôt pour être facilement exportable pour les services chargés de le fournir.

Enfin, est-ce que la transparence des marchés immobiliers est au rendez-vous de cet Open Data ?

Alexandre Coulondre : Cette question a été au cœur de travaux récents que j'ai menés avec Claire Juillard[6]. L'histoire nous apprend que l'ouverture des données et ses modalités pratiques sont le fruit de choix politiques. Le choix qui a été fait en France consiste à diffuser de la donnée brute. Ceci favorise le développement d'intermédiaires compétents capables de traiter cette donnée complexe pour le public. On aurait pu imaginer que l'Etat conserve une forme de centralisation avec une administration qui prétraite et met en forme la donnée pour que le grand public y accède sans autre intermédiaire.

Quel choix aurait permis plus de transparence ? C'est difficile à dire puisqu'on n'a qu'une version de l'histoire. Ce que l'on peut dire c'est que les travaux qui ont étudié les conséquences de l'Open Data tel qu'il est à l'œuvre aujourd'hui montrent que les ouvertures ne débouchent pas toujours sur une lecture plus limpide des marchés immobiliers par le grand public. Dans certains marchés, l'abondance de données produit plutôt de la confusion[7].

Entre la donnée et la connaissance il y a encore au moins une étape qui est celle de l'outil statistique. Or, ces outils statistiques ne sont pas neutres. Ils sont développés par des intermédiaires privés. Plus de 60 peuvent être identifiés[8]. La plupart ont connu un essor récent. Citons par exemple MeilleursAgents, Bien'ici, Bienestimer, ou encore Homeloop, etc. Ces entreprises proposent gratuitement des cartes de prix et des outils d'estimation automatique de la valeur des logements. Ces outils ne se fondent pas uniquement sur DVF mais sont largement nourris par l'Open Data et ont connu de nouveaux développements avec le mouvement d'ouverture des données publiques. Cette mise à disposition d'outils de connaissance des marchés s'inscrit dans un modèle économique qui est celui des plateformes. Il consiste à faire payer par ailleurs les professionnels de l'immobilier pour obtenir de la visibilité sur le site et collecter des contacts qualifiés. De fait, la donnée immobilière est donc devenue un marché.

En devenant un marché, l'information immobilière passe d'un problème d'accès à un problème d'« opacité méthodologique » pour reprendre l'expression de plusieurs géographes[9]. Les acteurs du marché sont nombreux. Ils n'utilisent pas forcément les mêmes sources de données. Beaucoup mélangent des données ouvertes et des données internes. Quand ils utilisent les mêmes sources, ils mettent en œuvre des méthodes différentes pour produire leurs indicateurs de prix. Or, ces choix méthodologiques ne sont pas vraiment expliqués pour des raisons de propriété industrielle. Ceci limite de fait la capacité du public à comprendre l'information assez variée sur les prix qu'il peut collecter auprès des intermédiaires.

L'Open Data déplace le débat sur la transparence. Il fait émerger un enjeu de transparence sur les méthodes, mais aussi un enjeu de création de référentiels et de normes d'utilisation partagées par les différents intermédiaires du secteur.

Parallèlement apparaissent de nouveaux types d'observatoires et de ressources mises en ligne par des organismes publics qui reposent sur d'autres modalités de collecte, non exhaustives et donnant lieu à des modélisations et redressements importants tels que l'observatoire GEODIP sur la précarité énergétique (ONPE-Ademe) ou la Carte des loyers mise en ligne par le ministère du logement en partenariat avec l'ANIL. Dans les deux cas, sont fournies des données communales pour l'ensemble du territoire. Qu'en pensez-vous ?

Alexandre Coulondre : Bien sûr, la tendance qui consiste à fournir de plus en plus de données brutes ne gomme pas le rôle de l'Etat dans la production d'expertises et de récits sur les territoires.

Pour comprendre cet apparent paradoxe, il faut préciser que les outils de connaissances auxquels vous faites allusion émanent de segments de l'Etat dont la mission est précisément de produire de la connaissance sur les territoires. Ce qui n'est pas le cas du principal fournisseur de données brutes en immobilier : la DGFIP. Dans le cas de l'ADEME pour les questions énergétiques, ou même de l'INSEE pour les questions démographiques, la cohérence statistique prévaut sur l'exhaustivité dans la mise à disposition des informations. Celles-ci sont restituées au terme de procédures de vérification, de gestion des valeurs manquantes et de redressements. On pourrait dire que ce qui est diffusé est plus proche de la connaissance que de la donnée brute. Ces deux modalités cohabitent au sein de l'Etat.

Vincent Lasserre-Bigorry : Il existe aussi de nombreuses sources encore extérieures au giron de l'Open Data. Dans ces cas-là, il est difficile d'aller plus loin qu'une diffusion d'indicateurs agrégés à la commune. C'est le cas de la Carte des Loyers dont la méthodologie s'appuie sur des données privées issues des plateformes d'annonces immobilières. Il en va de même pour l'Observatoire National de l'Artificialisation des Sols qui s'appuie notamment sur les Fichiers Fonciers, eux-aussi encore réservés à certains ayants droit publics. Une diffusion brute n'est pas envisageable ici pour des questions juridiques.

Propos recueillis par la rédaction de Politiquedulogement.com
Février 2023

[1] Alexandre Coulondre est chercheur associé au Lab'Urba (Université Gustave Eiffel), consultant indépendant (DIT Conseil), et animateur du comité « Data » du LIFTI (Laboratoire d'Initiatives Foncières et Territoriales Innovantes)

[2] Vincent Lasserre-Bigorry est chercheur associé au LVMT (Université Gustave Eiffel)

[3] Voir notamment : Boulay G., Buhot C., Fournier J.-L. (2012), « [Les chercheurs exclus de l'Open Data ? Appel à signatures lancé à la communauté de la recherche](#) », *Cybergeo : European Journal of Geography* [En ligne].

[4] Loi pour un État au service d'une société de confiance du 10 août 2018.

[5] Bléhaut Marianne, Coulondre Alexandre et Juillard Claire (2022), « [Un exode urbain post-covid ? Analyse des projections géographiques des Français à partir des données du site d'annonces immobilières leboncoin](#) », rapport pour POPSU et le Réseau Rural Français, septembre.

[6] Coulondre Alexandre avec Juillard Claire (2022) « [Le marché immobilier à l'ère de la donnée : une transparence accrue au service de nouveaux intermédiaires ?](#) », Working Paper de la Chaire Villes et Numérique de Sciences Po, n°01-22.

[7] 28% des plus grands consommateurs de données de prix au moment d'une vente considèrent que les informations récoltées ne sont pas concordantes, contre 18% chez ceux qui en consomment modérément (Coulondre et Juillard, p.16, *op. cit.*).

[8] (Coulondre et Juillard, *op. cit.*).

[9] Boulay, Guilhem, Delphine Blanke, Laure Casanova Enault, et Alexandre Granié. « [Moving from market opacity to methodological opacity : Are web data good enough for french property market monitoring ?](#) » *The Professional Geographer*, 2020, 1 16.