



L'ouverture des données foncières, une utopie du passé ?

Thibault Lecourt

► To cite this version:

Thibault Lecourt. L'ouverture des données foncières, une utopie du passé ?. Fonciers en débat, 2025. <hal-04926257>

HAL Id: hal-04926257

<https://hal.science/hal-04926257v1>

Submitted on 3 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.



Distributed under a Creative Commons CC BY-NC-SA 4.0 - Attribution - Non-commercial use - ShareAlike - International License

L'ouverture des données foncières, une utopie du passé ?

Thibault Lecourt, doctorant et A.T.E.R en Géographie & Aménagement

Avignon Université, UMR ESPACE – Chaire GIF

thibault.lecourt@univ-avignon.fr

« Seuls, les universitaires et quelques fonctionnaires dévoués qui ont le courage de se plonger dans les couches successives de documents accumulés dans les archives des mairies et des préfectures, ont quelque chance d'aboutir à des descriptions intéressantes, à condition encore qu'ils restreignent le champ de leur investigation à une petite région »

Edgar Pisani, *Utopie foncière*, 1977

Ces mots d'Edgar Pisani nous montrent tout le chemin qui a été parcouru pour l'ouverture des données foncières depuis 50 ans. À une époque où l'idée d'un grand « livre foncier » recensant l'ensemble des informations relatives à la propriété foncière était une utopie, cette ambition reposait sur un objectif fort : lutter contre la spéculation, protéger les terres agricoles et naturelles, et garantir une gestion publique juste et rationnelle du sol. Aujourd'hui, l'accès aux données foncières n'a jamais été aussi facile, mais paradoxalement, l'idéal d'une gestion collective du foncier ne semble pas être un horizon désirable pour les pouvoirs publics. Dès lors, quelle nouvelle utopie est sous-tendue par la massification contemporaine des données foncières ? C'est à cette question que s'est attachée l'atelier 3A du colloque « Pour une nouvelle utopie foncière », croisant les regards de la recherche avec des approches opérationnelles : Aline Clozel, responsable du service Habitat (Grand Avignon), Vincent Cézard, responsable du pôle Études et Développement (EPF 74) et Maud Anaïs Claudot, Fondatrice et CPO de l'entreprise Buildrz. Ce texte se situe dans son prolongement et n'engage que son auteur.

La massification des données foncières, levier ou verrou ?

L'ouverture et l'accessibilité des données foncières est une réalité incontestable aujourd'hui, mais une réalité à géométrie variable. Les listes des données disponibles, avec leurs spécificités, leurs atouts et leurs limites, ne manquent pas (Coulondre, 2018). La plupart restent en accès limité : soit payantes (Bien, Perval), soit gratuites mais restreintes à une zone géographique, une temporalité et/ou un type d'acteur ayant-droit. C'est notamment le cas de la base de données réputée la plus exhaustive en France aujourd'hui sur la propriété foncière bâtie et non bâtie : les Fichiers Fonciers, issus du retraitement et de l'enrichissement des données d'origine fiscale par la DGALN et le CEREMA. Cette ouverture est d'abord effectuée au bénéfice des collectivités locales : selon le mantra « observer pour agir » plébiscité par Aline Clozel, elle vise à doter les acteurs locaux des informations essentielles pour planifier l'aménagement du territoire en objectivant les besoins, les ressources et les opportunités. Le monde de la recherche universitaire est quant à lui resté un temps à l'écart de cette ouverture, jusqu'à ce que la mobilisation des équipes de recherche concernées porte ses fruits (Boulay et al., 2012). Si les acteurs privés ne figurent pas, à ce jour, parmi les ayants-droits des Fichiers Fonciers, certaines données sont depuis peu disponibles en *open data* ([DVF, Locaux et Parcelles des Personnes Morales](#)), publiées en licence ouverte ETALAB autorisant notamment leur exploitation à des fins commerciales. En d'autres termes, l'ouverture récente et progressive des données foncières répond indéniablement aux besoins des pouvoirs publics et de la recherche, et ouvre aussi désormais des opportunités pour les acteurs privés.

Néanmoins, le foisonnement de données foncières n'est pas sans générer en même temps une certaine confusion. Données de transaction d'origine fiscale, données transmises par les notaires, données sur la propriété foncière, sur les permis de construire, sur la production de logement... Il est parfois difficile de s'y retrouver dans cet ensemble, dont l'objet, la structure, les variables, les couvertures spatiales et temporelles varient d'une base à l'autre. Une hétérogénéité qui peut être un frein à l'exploitation des données foncières, rendant les croisements et les comparaisons entre les sources de données parfois périlleux. Il a par exemple été montré qu'une transaction ne revêt pas la même définition selon l'approche de la DGFIP ou celle de la base Perval¹, mettant même en avant « l'hétérogénéité des modes d'interprétation de certaines variables » dans cette dernière (Casanova Enault et al., 2017).

En parallèle de leur hétérogénéité, les différentes sources de données foncières font aussi preuve d'une certaine instabilité dans les variables qui les composent. L'enrichissement permanent apporté par les producteurs de données répond à la multiplicité croissante des données disponibles, des capacités de traitement, des publics ayant-droit et donc des thématiques à couvrir, mais induit paradoxalement des difficultés supplémentaires dans la mobilisation de ces données. Les statistiques communales sur les logements vacants produites dans la [base LOVAC](#) par le CEREMA ont par exemple été redéfinies en 2021, contraignant les collectivités qui l'utilisaient à revoir certaines de leurs procédures. Faut-il une amélioration continue des données produites afin d'être plus précis,

¹ La base DVF, dont la constitution poursuit un but de prélèvement de l'impôt, considère une transaction entre deux parties portant sur plusieurs biens (appartement et cave, maison et terrain agricole attenant...) comme une seule mutation comprenant plusieurs enregistrements : l'identifiant de la transaction se répète sur plusieurs lignes. A l'inverse, Perval identifie quel est le bien principal concerné par la transaction pour n'en retenir qu'une ligne, les annexes étant consignées dans une colonne dédiée.

plus complet ; ou faut-il une stabilisation des données disponibles pour favoriser leur compréhension et leur utilisation dans le temps ? Difficile de répondre à ce dilemme sans se (re)poser la première question qui doit être posée : à quels besoins répondent les données foncières ?

Se reposer la question des besoins : des données foncières pour quoi faire ?

Les données portent très mal leur nom : elles ne sont jamais « données » : elles sont à saisir, à comprendre et à contextualiser pour ensuite pouvoir être correctement interprétées en fonction de son besoin. Quel sens aurait de calculer un prix moyen de l'immobilier dans une commune sans distinguer des sous-marchés ? Le prix au m² d'un T2 neuf ne peut être comparé à celui d'une ancienne maison de ville, tout comme le prix d'une terre agricole ne doit pas être confondu à celui d'un terrain à bâtir. Dès lors, les données sur les transactions foncières doivent être appréhendées avec un regard métier expert capable de les disséquer pour en tirer la « substantifique moelle », en fonction de ce que l'on cherche. L'enjeu est donc de rendre intelligibles les données foncières, pour répondre à des questions et des objectifs d'aménagement.

A cette expertise métier, nécessaire pour se poser les bonnes questions, se superpose un besoin d'expertise technique des données foncières. Ces dernières, encore une fois, ne sont pas « données » mais accessibles via des sources différentes, selon des licences d'utilisation différentes, dans des formats différents et avec des structures de base de données différentes. Dès lors, les interactions nécessaires entre données foncières impliquent des compétences de déploiement, de gestion et d'interrogation de bases de données, mais aussi de sécurisation dans le cas des données sensibles à diffusion limitée. En fonction de l'échelle géographique à laquelle on travaille, la question du volume des données se pose très rapidement, et nécessite une équipe capable de gérer son stockage et ses performances d'interrogation. Ces contraintes techniques sont parfois de vrais freins à l'utilisation efficace, ou à l'utilisation tout court, des données foncières par les acteurs publics comme privés, dans le monde opérationnel comme dans celui de la recherche. L'expertise technique nécessaire peut ainsi parfois prendre le pas sur les besoins d'analyse réellement identifiés, en précipitant la question « Qu'est-il possible de dire avec ces données » avant celle, pourtant fondamentale : « Qu'a-t-on besoin de savoir ? ».

Conjuguer expertise technique et expertise métier apparaît donc indispensable pour manipuler les données foncières, mais cette association n'est pas toujours aisée à mettre en œuvre dans des collectivités locales aux ressources limitées. Afin de pallier ce problème, plusieurs Établissements Publics Fonciers (EPF) jouent le rôle de facilitateur dans l'exploitation des données foncières par leurs collectivités membres. Au-delà de leur rôle central dans le portage foncier, ces établissements sont en effet aussi moteurs dans le conseil expert sur les stratégies et les politiques foncières, et dans le diagnostic des besoins et des ressources. Certains EPF, comme celui de Haute-Savoie sont allés jusqu'à mettre en place des systèmes d'information géographique en ligne² pour permettre aux collectivités de visualiser et mettre à jour facilement un grand nombre de données foncières sur leur territoire ; une démarche similaire à celle portée par le CEREMA à l'échelle nationale avec l'outil

² L'EPF74 a mis en place l'outil appelé « Mieux connaître pour mieux agir », [MCMA](#)

[Urban Simul](#). La co-existence de ces outils pose question : ceux développés localement présentent l'immense avantage de pouvoir s'adapter facilement aux besoins et ressources identifiés par les utilisateurs, et d'intégrer des données locales au système produites spécifiquement par une commune (un diagnostic des logements vacants réalisé sur le terrain, par exemple). Mais ils présentent l'inconvénient de n'être disponibles que sur les territoires où l'EPF a mis en place de genre d'outil, ce qui exclut de fait tout une partie du territoire, et qui complique la tâche des agglomérations dont le territoire est couvert par plusieurs EPF comme le Grand Avignon, entre l'EPF PACA et l'EPF Occitanie. Une plateforme à l'échelle nationale revêt ainsi l'avantage de l'harmonisation des données sur tout le territoire, au risque d'être parfois inadaptée.

Imperfections et incertitudes : les limites intrinsèques des données

En effet, les données ne sont jamais neutres, et les données foncières n'y échappent pas. Elles sont toujours produites dans un contexte, par des institutions répondant à leurs objectifs propres. Les travaux de Mathieu Noucher (2023) ont bien montré que le foisonnement de données géographiques à l'ère de l'*open data* ne doit pas faire oublier l'éternel « blanc des cartes », ces espaces et ces informations oubliés, volontairement ou non. Les données foncières sont nécessairement lacunaires, elles reflètent le regard que portent leurs producteurs sur le foncier, et ne sont donc pas adaptées à tous les usages (Ouvrard et Guéringer, 2023). Le CEREMA en donne un exemple éclairant dans son [guide d'utilisation des Fichiers Fonciers](#) : le type d'occupation du sol des parcelles est déclaré par les contribuables dans le cadre du prélèvement de la taxe foncière par les services des impôts ; dans la mesure où certains changements d'occupation du sol n'engendrent pas de changement d'impôt, ils ne sont pas nécessairement déclarés ni vérifiés par la DGFIP, dont la mission n'est pas de constituer une base de données exhaustive et fiable de l'occupation du sol, mais de prélever l'impôt. Utiliser les données d'occupation du sol dans les Fichiers Fonciers appelle à beaucoup de vigilance pour s'assurer qu'elles permettent bien de répondre à nos besoins. C'est pourquoi il est fondamental que les métadonnées informent systématiquement sur le contexte et explicitent les choix méthodologiques qui ont orienté leur production.

L'enjeu aujourd'hui n'est donc plus tant dans l'ouverture des données foncières que dans la compréhension de leur portée et de leurs limites. En particulier, force est de constater que l'essentiel des données actuellement disponibles informent principalement sur une dimension bien spécifique du foncier : sa dimension économique. Prix des transactions, valeur locative, estimation de la valeur foncière, taux d'impositions... L'analyse de ces bases amène naturellement à considérer le foncier comme un bien marchand, et passe largement sous silence d'autres approches plus écosystémiques. Des bases de données sur la qualité des sols et des sous-sols, sur la biodiversité, sur la valeur d'usage et l'intérêt social du foncier, peuvent exister localement mais sont bien plus rares. Ce constat n'enlève rien à l'intérêt des bases de données portant sur la dimension économique du foncier, mais amène d'une part à relativiser les analyses qui s'appuient sur elles, et d'autre part à interroger les usages à venir de l'analyse automatisée des données par intelligence artificielle.

Carte blanche à l'intelligence artificielle ?

Le développement récent et rapide de l'IA n'épargne évidemment pas les données foncières, et ouvre des perspectives prometteuses. Si la modélisation automatisée de bâtiments respectant les règles de constructibilité locale est déjà une réalité à l'échelle d'une parcelle, on peut facilement imaginer que d'ici peu, il suffira de quelques clics pour visualiser en temps réel l'impact sur la forme urbaine de l'évolution d'une règle dans le PLU. Les travaux et réflexions de Buildrz sont en ce sens une avancée majeure dans l'aide à la décision des acteurs publics, avec l'immense avantage de les affranchir de la coûteuse expertise technique aujourd'hui nécessaire à l'exploitation des données foncières nombreuses, complexes et hétérogènes. L'assistance de l'IA dans ce domaine reste néanmoins limité, et l'intervention humaine nécessaire, lorsqu'on travaille avec la réglementation. Les services numériques de modélisation de potentiel bâti selon la réglementation locale sont déjà une réalité, ils reposent sur des algorithmes calibrés selon les règles du PLU ; mais la transformation en langage binaire de règles d'urbanisme est une tâche qui relève parfois plus de l'interprétation que de l'application de contraintes clairement définies, si bien que l'appel au service urbanisme de la collectivité est parfois nécessaire pour saisir l'« esprit » de la règle. Et c'est sans compter les cas où la règle est peut-être volontairement floue, permettant à certaines collectivités de choisir au cas par cas les projets qu'elle autorise. L'intelligence artificielle n'est donc pas nécessairement toujours une réponse adaptée pour faciliter une opération d'aménagement.

Par ailleurs, la perspective du développement de l'IA dans le traitement des données foncières fait courir le risque de masquer, derrière une certaine démocratisation de l'utilisation des données, une opacité méthodologique préjudiciable à leur compréhension et leur interprétation. Autrement dit, le risque de remplacer la dictature de l'expertise technique par une dictature de la boîte noire algorithmique. La capacité des algorithmes d'apprentissage automatique à produire des « super-indicateurs » croisant un grand nombre de données pourrait faire oublier que les données utilisées ne sont pas neutres. Loukissas (2019) a par exemple montré comment la plateforme d'annonces immobilières Zillow aux Etats-Unis, qui revendique une analyse objective des marchés basée sur des données immobilières transparentes, influence en réalité la perception du marché local par les utilisateurs, et produit des effets sur ce marché, de l'ordre de l'auto-prophétie. En d'autres termes, la prédiction présumée objective des prix parce qu'appuyée sur des données publiques, n'est pas sans impact sur... les prix eux-mêmes.

Si l'« utopie foncière » n'est plus à l'ouverture des données, on peut sans doute souhaiter qu'elle repose sur un usage maîtrisé et réfléchi de ces données, afin de répondre aux enjeux de sobriété et de justice foncières qu'avait bien anticipés Edgar Pisani.

Références

Guilhem Boulay, Clotilde Buhot, Fournier Jean-Louis, « Les chercheurs exclus de l'open data ? Appel à signatures lancé à la communauté de la recherche », *Cybergeo : Revue européenne de géographie / European journal of geography*, 2012, DOI : 10.4000/cybergeo.25520. {10.4000/cybergeo.25520}. {hal-01191849}

Laure Casanova Enault, Guilhem Boulay, Yann Gérard, Lisa Yahi, « Deux bases de données, aucune référence de prix. Comment observer les prix immobiliers en France avec DVF et Perval ? », *Revue d'économie régionale et urbaine*, 2017. {hal-01620667}

Alexandre Coulondre. « Ouvrir la boîte noire des marchés du logement », *Métropolitiques*, 2018. {hal-01838393}

Yanni Alexander Loukissas, *All Data Are Local: Thinking Critically in a Data-Driven Society*, The MIT Press, 2019, DOI : <https://doi.org/10.7551/mitpress/11543.001.0001>

Matthieu Noucher. *Blancs des cartes et boîtes noires algorithmiques*. CNRS Editions, 2023. {halshs-04207881}

Antoine Ouvrard et Alain Guéringer, « La raison de la donnée est toujours la meilleure : de l'utilisation des fichiers fonciers pour quantifier les dynamiques spatiales », *Économie rurale*, n°383, 2023, DOI : <https://doi.org/10.4000/economierurale.11165>