



HAL
open science

Plan de gestion de données du projet VALO : La vente HLM dans les régions lyonnaise et grenobloise : quelles valorisations?

Loïc Bonneval

► To cite this version:

Loïc Bonneval. Plan de gestion de données du projet VALO : La vente HLM dans les régions lyonnaise et grenobloise : quelles valorisations?. Université lumière Lyon 2; Institut d'urbanisme de Grenoble, 14 et 14 bis avenue Marie Reynoard, 38100 Grenoble, France. 2021. halshs-04668531

HAL Id: halshs-04668531

<https://shs.hal.science/halshs-04668531v1>

Submitted on 6 Aug 2024

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.

PROJET VALO LA VENTE HLM DANS LES REGIONS LYONNAISE ET GRENOBLOISE : QUELLES VALORISATIONS ?

PLAN DE GESTION DES DONNEES

Le projet intitulé *La vente HLM dans les régions lyonnaise et grenobloise : quelles valorisations ?* (en abrégé VALO) est financé par l'Union Sociale de l'Habitat (USH) dans le cadre du Programme de recherche USH Vente HLM 2020-23. Il propose de recenser les ventes HLM dans différents types de territoires de la région Auvergne Rhône Alpes, plus précisément dans les régions lyonnaise et grenobloise, dans des contextes de marché tendus et détendus, métropolitains et non métropolitains. L'objectif est de questionner les différentes facettes de la valorisation attendue et observée par la vente de logements HLM : valorisation économique avec l'observation des effets de ces ventes sur les marchés locaux du logement ; valorisation patrimoniale avec les stratégies des bailleurs et collectivités en la matière, le devenir du patrimoine immobilier des copropriétés mixtes ainsi créées, mais aussi des ménages ; valorisation territoriale enfin, notamment dans les quartiers en renouvellement urbain. Une analyse quantitative économétrique sera menée, au niveau national et dans les territoires d'étude en croisant les prix des ventes HLM avec ceux des marchés locaux. Des données plus fines et qualitatives seront recueillies grâce aux ressources des partenaires du projet, auprès des acteurs territoriaux et des ménages concernés, et permettront d'interpréter et/ou nuancer les résultats de l'analyse quantitative. Ce projet s'appuie sur un panel éprouvé de partenaires diversifiés.

1. DESCRIPTION DES DONNEES COLLECTEES OU REUTILISATION DE DONNEES EXISTANTES

1a. Comment de nouvelles données seront-elles recueillies ou produites et/ou comment des données préexistantes seront-elles réutilisées ?

Les données traitées dans le projet VALO proviennent à la fois de bases de données préexistantes et de données produites dans le cadre de la recherche.

Bases de données issues de la statistique publique, obtenues sur demande aux organismes diffuseurs (chacune des bases utilisées dispose de métadonnées et dictionnaires des variables) :

-Le **répertoire des logements locatifs des bailleurs sociaux (RPLS)**. Ces données seront utilisées pour recenser et caractériser les logements HLM vendus à des particuliers ainsi que leurs spécificités par rapport à l'ensemble du parc de logements sociaux et aux évolutions de celui-ci. Le Service des Etudes statistiques (SDES) du Ministère de la cohésion des territoires est chargé de la tenue du RPLS et de sa diffusion¹. Une partie est accessible en libre accès via l'open data des données publiques mais certaines variables nécessaires pour l'étude, par exemple celles permettant d'identifier l'organisme qui vend le logement, ne figurent pas dans le jeu de données en open data. Nous avons donc fait la demande, et obtenu, un accès au fichier complet pour l'ensemble de la France de 2011 à 2020.

¹ <https://www.statistiques.developpement-durable.gouv.fr/repertoire-des-logements-locatifs-des-bailleurs-sociaux-rpls> consulté le 01/06/2021

-La **base de données DV3F** est issue du croisement de la base DVF (pour « demandes de valeurs foncières » qui recense les mutations à titre onéreux publiées par les services de la publicité foncière sur l'ensemble du territoire français) produite par la DGFiP et des fichiers fonciers gérés par le Cerema. Elle constitue un outil pour étudier les marchés immobiliers à partir des données de la statistique publique. Elle sera utilisée pour décrire les contextes des marchés du logement dans lesquels sont vendus des logements HLM. Par ailleurs, dans la mesure où une variable indique si le vendeur est un organisme HLM, elle peut être comparée au RPLS pour s'assurer avoir un autre décompte des ventes HLM. DV3F est prise en charge par le CEREMA² mais est également détenue par la DGALN. Il existe une version de la base DVF en open-data (DVF, pour « demandes de valeurs foncières ») mais, là aussi, certaines variables nécessaires à la recherche ne figurent que dans DV3F. La possibilité de suivre l'évolution du prix de chaque logement au fil des ventes avec cette base est actuellement à l'étude. Nous avons donc adressé une demande au CEREMA et obtenu la base DV3F pour 2010-2019. Trois chercheurs de l'équipe ont fait une demande, nécessitant un acte d'engagement de leur laboratoire : Florence Goffette-Nagot pour le GATE (France entière), Loïc Bonneval pour le Centre Max Weber et François Robert pour Triangle (départements de l'Ain, du Rhône, de la Loire et de l'Isère).

Des demandes pour le RPLS vont également être faites par les deux autres laboratoires impliqués dans la recherche, PACTE et EVS.

Pour ces bases, les autorisations obtenues ne permettent des traitements que dans le cadre de la recherche VALO.

Les données produites dans le cadre de l'enquête

-les **données de la publicité foncière** : La recherche nécessite d'une part de remonter au-delà de 2010 et de voir si le prix des logements HLM vendus évolue lorsqu'ils sont revendus par le premier acquéreur. Or les actes de vente conservés par les hypothèques contiennent un historique des prix du bien. Les obtenir suppose toutefois de faire une demande aux services des hypothèques en deux temps : d'abord pour obtenir les ventes réalisées par chaque organisme, puis, avec les références des biens, pour chaque bien. Les ventes des 5 plus grands bailleurs de la métropole lyonnaise dans les secteurs correspondant aux bureaux des hypothèques couvrant la métropole sont demandées. Ces données peuvent être librement obtenues (c'est le principe de la publicité foncière) mais elles sont payantes et ne sont pas structurées en base de données. Ce travail nécessite donc de la saisie et la construction d'une base de données.

-Les **entretiens auprès des acteurs institutionnels**, bailleurs, collectivités locales, et organismes interbailleurs, visent à appréhender les logiques de la vente HLM, de la mise en place des plans de ventes au suivi des copropriétés après la vente, en passant par l'ensemble du processus de commercialisation, et à situer ces logiques dans les enjeux locaux des politiques de l'habitat. En amont de l'appel à projets lancé par l'USH, un certain nombre d'acteurs ont manifesté leur intérêt pour participer à cette recherche, et ils fournissent un premier vivier de contacts. Néanmoins, les entretiens vont au-delà et seront réalisés sur l'ensemble des terrains de la

² <https://datafoncier.cerema.fr/donnees/donnee-dv3f> consulté le 01/06/2021

recherche (cf. tableau ci-dessous). La participation de la société Espacité au projet Valo a pour objectif de faciliter les contacts avec ces acteurs institutionnels.

Lors des visites chez les organismes bailleurs ou les collectivités, des **documents** internes pourront être recueillis et exploités avec l'accord des intéressés. Il s'agit de documents ayant trait aux politiques et stratégies de ventes de logements sociaux : archives, de rapports internes, de correspondance, des données sur le patrimoine de l'organisme, des plans de vente, des délibérations sur les ventes réalisées, etc., de tout document écrit ayant trait aux politiques et stratégies de ventes de logements sociaux, de tout document écrit.

-Les **entretiens auprès d'habitants**, locataires et propriétaires : la recherche prévoit une campagne d'entretiens auprès de ménages ayant acheté leur logement après la mise en vente par le bailleur social, de nouveaux arrivants dans la résidence, ou encore de locataires HLM dans des résidences ayant fait l'objet de telles mises en vente. Les résidences étudiées seront choisies parmi les terrains d'études listés ci-dessous. L'accès aux personnes enquêtées se fera en grande partie par l'intermédiaire du bailleur : celui-ci réalisera une information auprès des résidents (par exemple sous forme de mot dans les boîtes aux lettres ou affiché dans le hall) et communiquera aux enquêteurs leurs coordonnées afin qu'ils puissent les contacter directement. Les résidents pourront choisir d'accepter ou non l'entretien. Dans certains cas, le bailleur pourra être amené à se limiter à l'information aux habitants, et les enquêteurs iront à leur rencontre au porte-à-porte pour solliciter leur consentement pour l'entretien. Les coordonnées ne seront utilisées que pour la prise de contact et ne donneront pas lieu à l'établissement d'une base de données.

Dans les deux cas, les entretiens sont enregistrés et retranscrits intégralement. Chaque retranscription inclut un bref descriptif (date, lieu, durée, précisions sur l'enquêté et fonction s'il appartient à une institution en lien avec la vente HLM). Un tableau récapitulatif des entretiens permet de recenser tous ceux réalisés.

Les terrains de la recherche Valo

Département	EPCI	Commune
Rhône	Métropole de Lyon	Lyon (tous ards)
		Bron
		Villeurbanne
		Neuville sur S.
		Vénissieux
		Grigny
		Saint-Fons
		Ecully
		Saint-Genis
		Monts du Lyonnais
	Villefranche-Beaujolais	Limas
	Pays de Tarare	Tarare
Loire	Roannais agglomération	Roanne
Isère	CAPI	Villefontaine
		Isle-d'Abeau
		Bourgoin-Jallieu
		Saint-Quentin-Fallavier
	Grenoble Alpes Métropole	Grenoble
		Saint-Martin-d'Hères
		Pont-de-Claix
	CC du Grésivaudan	Crolles
		Villard-Bonnot
	CC Saint-Marcellin Vercors Isère Communauté	Rencurel
		Pont-en-Royans
	Vienne Condrieu Agglomération	Chasse-sur-Rhône
		Condrieu
Communauté du Pays Voironais	Voreppe	

1b. Quelles données (types, formats et volumes par ex.) seront collectées ou produites ?

Pour les bases de données existantes

–RPLS : les données sont fournies en format .csv. Il y a un fichier par année. Chaque fichier comporte environ 5 millions de lignes (3 Go chacun), une ligne par logement social, et une centaine de variables. Ils sont traités avec le logiciel R. Des extractions, avec parfois des compléments pour les données manquantes (par exemple le libellé de l'EPCI, LIBEPCI, quand on n'a que le code) ne comportant que les ventes HLM ont été réalisées. Dans ces fichiers, les ventes sont identifiées par la variable SORTIEPATRIM, modalités 1 (vente à l'occupant) et 3 (autre vente), cette dernière étant couplée avec la variable QUALACQ lorsqu'elle est disponible (modalités 1 à 5 c'est-à-dire les ventes à d'autres personnes physiques). Les ventes en bloc, à

d'autres bailleurs ou à des collectivités locales n'entrent pas dans le champ de l'étude. Ces extractions se présentent sous forme de fichiers excel (1 fichier par département concerné par l'étude « Vente Rhône.xlsx » « Vente isère.xlsx » « Vente Loire.xlsx ») : ce format est facilement utilisable par l'ensemble des membres de l'équipe de recherche et permet de disposer d'un outil de travail commun pour recenser les ventes et identifier les terrains sur lesquels mener les enquêtes qualitatives. Elles contiennent des variables texte, numérique et d'information géographique (du code de la région à celui de l'IRIS ainsi que l'adresse en toutes lettres) et les coordonnées XY en Lambert 93.

-DV3F compte 13 tables dont la principale (« mutations ») d'environ 300 000 lignes pour le seul département du Rhône. 130 variables environ sont renseignées. Les fichiers compressés pèsent environ 400 Mo. Pour leur exploitation, le fournisseur de données demande de disposer d'un environnement Post Gre SQL ainsi que de Post GIS pour le traitement des données localisées. Les premiers traitements, et notamment les requêtes permettant d'extraire les ventes par des organismes HLM ont été réalisés avec le logiciel R. Elle contient des variables texte, numérique et d'information géographique (les coordonnées XY en Lambert 93).

Pour les données produites dans le cadre de l'enquête

-Les données de la publicité foncière peuvent être saisies dans un tableur. Elles comportent pour chaque vente la date de l'acte, le prix, la localisation du bien à la parcelle cadastrale et le n° de lot de copropriété, le type de droits cédés (le plus souvent pleine propriété), la raison sociale du vendeur, le nom et la date de naissance de l'acheteur (le nom n'étant pas saisi ni conservé), la répartition des droits entre acheteurs dans le cas d'un achat commun, les références d'enlissement de l'acte de vente.

-Les entretiens seront enregistrés et retranscrits respectivement en format audio (type mp3) et dans des traitements de texte usuels (Word ou open office). Il est envisagé environ 50 entretiens avec des acteurs institutionnels et 50 avec des habitants. Une fiche de synthèse par entretien sera incluse dans le fichier de la retranscription. La convention de nommage est la suivante : institution_date_fonction de la personne interrogée.

2b. Quelles mesures de contrôle de la qualité des données seront mises en œuvre ?

En ce qui concerne les bases de données existantes, qui entrent dans le champ de la statistique publique, les procédures de collecte et d'établissement des bases sont connues et communiquées par les diffuseurs. Les bases RPLS et DV3F sont censées être exhaustives. Elles nécessitent toutefois un travail de nettoyage et de contrôle des données manquantes. Celui-ci est réalisé en parallèle par deux groupes de travail, l'un travaillant à l'échelon national et l'autre à l'échelon local sur plusieurs terrains de la région Rhône-Alpes (cf. ci-dessus la liste des terrains). Des points réguliers entre ces groupes de travail permettent de repérer les incohérences dans les données et d'alerter sur les précautions à observer dans les traitements.

Un des enjeux de la recherche, le recensement des ventes HLM, suppose de confronter les bases RPLS et DV3F pour s'assurer que leurs résultats sont identiques, ou au moins convergents. L'objectif n'est pas de les appairer mais de vérifier la cohérence des données fournies par des sources différentes (l'une basée sur les déclarations des organismes HLM, l'autre sur des données fiscales). Un recoupement spatial, sur la base de SIG, pourra faciliter ce travail.

De la même façon, les données saisies à partir des registres de formalités de la publicité foncière pourront être comparées à ces bases pour voir si l'on retrouve dans chacune les mêmes transactions.

Enfin, les discours recueillis auprès des acteurs à propos de ces ventes permettront d'apporter des éclairages complémentaires permettant d'enrichir l'interprétation des données.

3. STOCKAGE ET SAUVEGARDE PENDANT LE PROCESSUS DE RECHERCHE

3a. Comment les données et les métadonnées seront-elles stockées et sauvegardées tout au long du processus de recherche ?

Les documents de travail, données et métadonnées sur lesquels travaillent plusieurs membres de l'équipe sont stockés sur le système de partage de fichiers³ de l'Université Lumière Lyon 2 (BUL) adapté au travail collaboratif et d'accès limité aux membres de l'équipe ayant accès aux données. Ils vont être transférés sur le serveur CNRS MyCore, mieux sécurisé. Il s'agit en particulier des extractions des bases RPLS et DV3F (limitées aux logements HLM vendus) ainsi qu'au fichier de saisie des données de la publicité foncière et des retranscriptions anonymisées d'entretiens. Les membres de l'équipe de recherche se sont engagés à ne pas diffuser les données à des tiers, ni les données brutes ni des données agrégées portant sur un effectif inférieur à 11 logements.

Les bases de données brutes (RPLS et DV3F) sont stockées sur un serveur de la TGIR Humanum⁴. L'accès et le traitement se font directement par le logiciel R.

3b. Comment la sécurité des données et la protection des données sensibles seront-elles assurées tout au long du processus de recherche ?

Les fichiers déposés sur l'espace de partage de fichier de Lyon II sont à la disposition de l'ensemble des membres de l'équipe et sont pensés comme des documents de travail commun.

Les données issues des bases de données de la statistique publique et de la publicité foncière sont localisées (coordonnées XY, adresse ou quartier IRIS) mais ne sont pas sensibles au sens de la CNIL : il s'agit principalement d'informations sur les logements et sur les données économiques (prix, loyer). Les variables sur les personnes portent sur la qualification (personne physique/morale), l'âge, le sexe. La réglementation à laquelle sont soumises ces bases est explicitée par les fournisseurs de ces données. Ces données ne peuvent pas être diffusées sous leur forme brute, elles doivent être agrégées (par groupes comptant au minimum 11 logements).

Les données nominatives sur les habitants ne seront communiquées qu'aux enquêteurs et uniquement à des fins de prise de contact en vue d'un entretien. Elles ne seront pas conservées

³ Cet espace de stockage est basé sur le logiciel OwnCloud

⁴ <https://www.huma-num.fr/>

ni communiquées à l'ensemble des membres de l'équipe. Les modalités de transmission de ces coordonnées sont à établir avec le bailleur qui les fournit.

4. EXIGENCES LEGALES ET ETHIQUES, CODES DE CONDUITE

4a. Si des données à caractère personnel sont traitées, comment le respect des dispositions de la législation sur les données à caractère personnel et sur la sécurité des données sera-t-il assuré ? Comment les éventuelles questions éthiques seront-elles prises en compte, les codes déontologiques respectés ?

En ce qui concerne les entretiens, il faut distinguer :

-Les entretiens avec des acteurs institutionnels : la question de l'anonymisation se pose principalement au moment de la restitution des résultats. Les propos ne seront pas attribués à l'enquêté de façon nominative, sauf consentement explicite fourni au moment de l'entretien, mais en présentant de façon large sa situation (fonction, type d'organisme, secteur géographique au sens large, c'est-à-dire au niveau de l'EPCI.

-Les entretiens avec des habitants seront anonymisés : aucune information nominative n'est conservée au-delà de la prise de contact et de la réalisation de l'entretien avec l'enquêté. Les personnes sollicitées sont informées en amont de l'entretien (information faite par le bailleur et répétée par les enquêteurs) que leurs propos resteront anonymes et qu'ils ont, conformément au RGPD, un droit d'accès à leurs données. Dans la restitution de l'étude (valorisation, publications) un nom fictif est attribué à chaque enquêté pour des raisons de lisibilité mais il n'est pas établi de tableau de correspondance permettant de retrouver les enquêtés (qui ne seront donc pas réinterrogés une 2^e fois dans le cadre de cette enquête ou de ses prolongements).

4b. Comment les autres questions juridiques, comme la titularité ou les droits de propriété intellectuelle sur les données, seront-elles abordées ? Quelle est la législation applicable en la matière ?

La question de la propriété intellectuelle se pose pour les données produites dans le cadre de l'enquête, celles fournies par des organismes leur appartenant. Elles ne peuvent être réutilisées par des tiers.

Pendant la durée du projet les données produites sont librement accessibles à l'ensemble des membres de l'équipe de recherche qui s'engagent à ne pas les utiliser dans un autre cadre ni à les transmettre.

Les données, et notamment les entretiens, produites en lien avec un stage, mémoire de master, thèse, ou post-doctorat sont librement utilisables à cette fin. Chaque membre de l'équipe de recherche utilisant les données en vue d'une publication s'engage à en informer les autres membres.

5. PARTAGE DES DONNEES ET CONSERVATION A LONG TERME

5a. Comment et quand les données seront-elles partagées ? Y-a-t-il des restrictions au partage des données ou des raisons de définir un embargo ? Comment les données à conserver seront-elles sélectionnées et où seront-elles préservées sur le long terme (par ex. un entrepôt de données ou une archive) ?

Les données diffusées par des tiers (bases RPLS, DV3F, notariales) ne sont délivrées que dans le cadre du projet VALO et ne peuvent être réutilisées pour d'autres projets. Elles ne sont donc pas conservées au-delà de la période de réalisation du projet et de publication des résultats.

Les données produites dans le cadre de l'enquête pourront faire l'objet d'un dépôt, par exemple par le biais d'une convention avec l'ADISP. Les modalités et le choix d'un entrepôt de données seront décidés dans une phase ultérieure du projet. Le dépôt pourra intervenir après la publication des résultats de l'enquête dans des revues, dans un délai de deux ans après la fin du projet.

6. RESPONSABILITES ET RESSOURCES EN MATIERE DE GESTION DES DONNEES

6a. Qui (par exemple rôle, position et institution de rattachement) sera responsable de la gestion des données (c'est-à-dire le gestionnaire des données) ? Quelles seront les ressources (budget et temps alloués) dédiées à la gestion des données permettant de s'assurer que les données seront FAIR (Facile à trouver, Accessible, Interopérable, Réutilisable) ?

Le nettoyage des données et des bases existantes est assuré par deux personnes recrutées en stage (sur RPLS et DV3F) encadrées par un ingénieur de recherche et un membre permanent du projet de recherche. Le travail de construction de la base et de saisie des données de la publicité foncière est réalisé par deux chercheurs de l'équipe de recherche. Les entretiens sont réalisés, pour le Rhône, par une doctorante et une ingénieure d'études recrutée pour le projet. Pour l'Isère, les entretiens sont réalisés et analysés par les deux enseignantes-chercheuses du laboratoire Pacte et par un stagiaire recruté par ce laboratoire. D'autres membres de l'équipe seront amenés à contribuer à la réalisation et l'analyse des entretiens (la grille d'analyse devant être discutée collectivement).

Les groupes de travail utilisant les mêmes jeux de données se tiennent au courant de leurs avancées et des différents aspects méthodologiques rencontrés. Les documents de travail méthodologiques sont mis à disposition sur l'espace de partage des fichiers. Le responsable de la gestion des données (Loïc Bonneval) s'assure que les ressources ainsi partagées permettent un usage FAIR des données (notamment par la relecture des documents de travail, et le lien avec les différents membres de l'équipe impliqués dans l'utilisation des données). Il a en charge l'actualisation du PGD.



triangle
UMR 5206

