



0'00 : Présentations de Karine Onfroy et Elise Thouron

Bienvenue,

Je suis Karine Onfroy, statisticienne à Bordeaux Sciences Economiques (BSE pour les initiés), une Unité Mixte de Recherche du CNRS et de l'Université de Bordeaux.

Je ne sais pas pour vous ... mais le retour d'expérience, c'est toujours le moment qui, soit me rassure et j'avance ... ou à l'inverse qui me fait abandonner toutes initiatives.

Ce n'est pas grave, c'est aussi à ça que ça sert. Alors je me suis dit que ce serait une bonne idée d'aller à la rencontre des doctorants de mon laboratoire qui ont osé lever le secret statistique ...

Toc, toc, toc.

Bonjour Elise, je peux te déranger ...

Oui, avec plaisir.

Je suis Elise Thouron, je suis en thèse CIFRE en urbanisme et aménagement.

Je partage mon temps entre le laboratoire BSE et l'agence d'urbanisme Bordeaux Aquitaine.

Je m'intéresse aux trajectoires résidentielles dans le pôle bordelais.

0'57 : Qu'est-ce qu'une thèse CIFRE ?

Thèse CIFRE, tu peux nous en dire plus ?

Une thèse CIFRE est un type de financement où l'on est salarié d'une entreprise ; dans mon cas d'une association : l'agence d'urbanisme Bordeaux Aquitaine, pendant trois ans. Je passe 60% de mon temps à l'agence d'urbanisme et 40% au laboratoire. Donc, je suis encadrée à la fois par des directeurs universitaires, des personnes du laboratoire BSE et de l'agence d'urbanisme.

Et que veut dire CIFRE ?

Convention Industrielle de Formation par la Recherche

Tu peux donc faire une thèse CIFRE dans le cadre de recherches en sciences humaines et sociales ?

Oui, il se trouve que c'est un peu plus récent mais c'est justement un des enjeux de l'[ANRT](#) (de développer plus de thèses en SHS).

Et donc pas que dans l'industrie.

Exactement.

2'00 : Pourquoi Elise s'est-elle intéressée aux données utilisées sur le CASD ?

Si l'on revient un petit peu en arrière, comment en es-tu venue aux données utilisées sur le CASD ? Tu avais une idée des sources dont tu avais besoin au départ ?

Je me suis intéressée aux données disponibles du CASD en connaissant déjà la base de données FIDELI [Fichiers Démographiques sur les Logements et les Individus].

Dans mon cas parce que j'avais eu accès à des tableaux de données utilisés dans le cadre d'un partenariat avec l'INSEE Nouvelle-Aquitaine. Je me suis posée la question de l'utilité d'accéder directement à cette base de données dans le cadre de mes travaux de thèse. C'est là que je me suis intéressée à la plateforme du CASD et à toutes les démarches en lien avec cette plateforme pour pouvoir accéder à la base directement.

C'était à quel moment de ta thèse ?

Milieu, fin de la première année.

Parmi l'ensemble des sources de données diffusées via le CASD, tu as cherché directement les informations concernant la base FIDELI ?

Oui, je travaille sur des données de logement et de ménages – individus. FIDELI est une base de données qui présente un intérêt assez fort car elle permet de mettre en lien par un identifiant commun les individus et les logements, et de pouvoir assurer un suivi dans le temps. C'est assez inédit par rapport aux bases de données INSEE de recensement de la population où l'on a des stocks à des instants T mais pas vraiment de flux. C'est pour ça qu'il y avait un intérêt à accéder à cette base de données.

Tu disais que la finesse d'information dans la base de données source peut t'intéresser.

Oui, est-ce que je ne pourrais pas avoir des choses en plus que ce que j'ai eu dans les tableaux de données.

3'42 : Quelles sont les différentes étapes pour accéder à la base FIDELI ?

Tu as effectué une demande d'accès au CASD pour la base FIDELI. Peux-tu nous raconter les différentes étapes ?

Je me suis lancée dans la procédure pour avoir accès à la base de données, j'ai obtenu l'accord du service producteur, l'INSEE, qui m'a posé la question de l'intérêt d'avoir accès à cette base de données par rapport aux bases de données des logements déjà existantes.

Ils voulaient être sûrs de tes besoins.

Oui et donc cette question est importante pour ceux qui se lancent dans la démarche. Cela demande de s'être bien renseigné sur les bases de données qui existent déjà dans le thème ou la question de recherche et de comprendre la plus-value de cette base de données avec des données confidentielles pour pouvoir le justifier auprès du service producteur.

L'INSEE voulait savoir si tu avais une bonne connaissance de tout ce qui existait en matière de sources de données liées à des thématiques du logement.

C'est ça, et est-ce que c'était nécessaire d'avoir accès à ce type de données.

C'est une question plus pour te faire réfléchir.

Oui, c'est ça.

Pour t'orienter.

Exactement.

Tu as obtenu ensuite un avis du Comité du secret statistique sur ton dossier ?

Oui, j'ai obtenu l'accord et ça débloque ensuite les démarches.

On reçoit le contrat du projet où il y a le montant associé dont je n'avais pas conscience au moment de la demande. Du coup, dans la balance avantages-inconvénients d'accéder directement à la base de données alors que j'avais eu les tableaux de l'INSEE Nouvelle-Aquitaine, qu'eux ont toute une expertise, qu'ils ont mis en place des kits, qu'ils ont des liens aussi avec l'INSEE nationale, et que par le partenariat a levé des barrières à l'entrée dans la base de données, je me suis dit au niveau du temps d'accès à la base de données, de la comprendre, de la traiter et du coût, il n'était plus forcément intéressant de m'y investir par rapport au coût que cela demandait, c'était plusieurs milliers d'euros.

Un partenariat avait été lié entre l'agence d'urbanisme et l'INSEE pour avoir des données plus agrégées.

Oui plus agrégées que ce qu'il y a dans la base de données parce que l'INSEE, et c'est normal, est prudente quand elle transmet ses tableaux de données pour ne pas violer le secret statistique.

Tu avais des données où il n'y avait pas de secret statistique.

Oui, on a fait une typologie du territoire c'est-à-dire que l'on a segmenté le territoire de la métropole et de la Gironde pour ensuite avoir les flux de mobilité entre ces différents territoires. C'est une façon d'avoir des données un peu plus fines mais qui ne soient pas non plus « à la ligne » comme dans la base de données.

Tu n'as pas une observation égale à un ménage ou à un logement.

Au-delà du secret statistique : comment accéder à des données confidentielles issues de la statistique publique ? Page 3 | 6

7'22 : Quel type d'informations recherche Elise ?

Pour collecter des données pertinentes, tu as aussi mis en place une enquête par questionnaire. Dans le cadre de cette enquête tu souhaites disposer de quelles informations ?

Comme je le disais, j'ai eu les tableaux issus de la base de données FIDELI et cela a permis d'avoir des flux de déménagements, des flux résidentiels entre territoires dans la métropole. Sauf que la limite, à cause notamment du secret statistique, c'est que je ne savais pas du tout qui étaient ces gens qui déménagent, quels sont leurs revenus, quels sont leurs statuts sociaux-économiques, le ménage auquel ils appartiennent.

Donc, il me manquait leurs motifs de déménagement, il manquait vraiment de la profondeur dans ces données.

Tu avais uniquement des flux quantitatifs d'un point à un autre.

J'avais de grandes données de cadrage sur le nombre de déménagements, les flux entre parcs de logement. C'est très bien d'avoir ça comme données puisque jusqu'à présent on avait que des stocks, on savait combien il y avait de personnes dans le parc locatif privé en 2017 par exemple et combien y vivaient en 2018 mais je ne pouvais pas savoir que tant de personnes sont passées du parc social au parc privé entre 2016 et 2018 par exemple. En termes de demande et de trajectoire résidentielle, j'étais un peu restée sur ma faim notamment dans le cadre de mes travaux de thèse.

La seconde partie de la thèse était de mettre en place une enquête « Nouveaux voisins », c'est une base de données mise à disposition par la Poste. On loue une liste de personnes qui ont déménagé au cours des six derniers mois entre décembre 2021 et mai 2022. Ce n'est pas moi qui se chargeait de la mise sous pli et de l'envoi, je n'ai jamais vu les adresses, je n'ai aucune information sur qui recevait le courrier, juste que ce sont des personnes qui ont déménagé mais « juste que » c'est très important pour moi car ce sont des gens qui connaissent bien le marché de l'immobilier bordelais tel qu'il est aujourd'hui et qui ont une réflexion par rapport à leur parcours résidentiel et leur trajectoire. Et ces mobilités sont très intéressantes pour comprendre la demande dans la métropole.

9'42 : Un choix méthodologique et financier

En fait, d'un point de vue financier tu t'es trouvée face à un choix méthodologique ?

Exactement, il fallait choisir entre acheter la base de données FIDELI ou mettre en place l'enquête en louant la base de données « Nouveaux voisins ». Et ayant déjà eu accès à la base de données FIDELI par des tableaux que l'on a valorisés dans une publication (**Insee Analyses Nouvelle-Aquitaine, n° 114, Janvier 2022**)*, j'ai priorisé l'enquête « Nouveaux voisins » qui permettait d'avoir des données plus fines et d'avoir accès à l'ensemble du parcours résidentiel de la trajectoire et pas qu'au déménagement. Cela permettait aussi d'aller un peu plus en

amont et en aval du dernier déménagement. C'est tout l'intérêt de l'enquête « Nouveaux voisins » et c'est aussi par le biais de cette enquête que je pouvais avoir des contacts. A la fin du questionnaire, je proposais de participer à des entretiens semi-directifs. Cela m'a aussi permis de faire ces entretiens pour mieux comprendre et mettre en récit les trajectoires résidentielles.

C'est pour ça que cela a orienté ton choix budgétaire plus vers cette base de données ?

Exactement, on est dans les mêmes ordres de grandeur au niveau financier, c'est comme ça que j'ai tranché.

11'00 : Comment Elise a-t-elle géré ses demandes d'accès à des données personnelles ?

Que ça soit dans FIDELI ou dans ton enquête postale, tu travailles sur des données personnelles. J'aimerais savoir dans quel registre des activités de traitement tu as enregistré ton travail sur les données ?

Tu es encadré par deux directeurs universitaires qui sont dans des UMR CNRS distinctes mais aussi par des personnes de l'Agence d'Urbanisme.

Comment as-tu fait ?

J'ai fait le plus sûr. A la fois je me suis enregistrée à la CNIL par le biais du juriste de l'agence d'urbanisme où je travaille, qui a mis dans son registre cette enquête par questionnaire. J'avais fait aussi des démarches, dans le cadre de la demande potentielle au CASD qui finalement n'a pas aboutie. J'ai appris beaucoup de choses en échangeant avec le CNRS sur la protection des données où j'ai bien pris conscience des enjeux de la protection des données, de comment stocker ces données à la fois papier et en ligne. Tout ça est important de s'y intéresser en thèse et c'est vrai que ce travail d'aller à la rencontre du CNRS, de comprendre et de savoir comment rédiger mon appel à participation, cela m'a permis de réaliser tous les enjeux liés à la protection des données.

C'est vrai que tu as eu la chance de rencontrer la déléguée à la protection des données du CNRS qui est située à Pessac près de Bordeaux et au sein de l'agence d'urbanisme qui était chargée de la déclaration de traitement ?

On a aussi un DPO.

Tu as rencontré les deux personnes ressources.

Exactement.

12'41 : Dans quel cadre Elise a-t-elle eu connaissance du CASD ?

Est-ce que tu as entendu parler du CASD dans ton cadre professionnel ?

Déjà dans le cadre d'échanges avec d'autres doctorants qui eux aussi étaient dans cette démarche d'accéder à des données via le CASD.

Tu es devenue un peu une sorte de personne ressource grâce à ton retour d'expérience ?

Plutôt on en a discuté avec ceux qui étaient dans cette démarche d'accéder à une base de données. Donc, on se tient au courant d'où on en est, si l'on a poursuivi ou pas comme la procédure est longue et demande d'aller à Paris.

13'23 : Le message d'Elise aux futurs utilisateurs du CASD

Pour finir, un message court à faire passer à de futurs utilisateurs du CASD ?

Oui, il faut anticiper. C'est important mais c'est faisable. Donc être confiant sur l'accès à ces données. Si l'on a de bonnes raisons de vouloir accéder à cette base de données, il n'y a pas de raison que les producteurs bloquent. Il faut juste bien justifier.

Toute cette expérience montre que l'on peut accéder à des données autrement que via le CASD par un partenariat avec l'INSEE où l'on a accès à des données à un niveau plus agrégé.

Ce qui est intéressant dans ce type de démarche aussi c'est que l'on a un lien sur le site du CASD où l'on apprend autre chose sur la base de données : j'ai pu voir les différentes variables. Cela me permettait d'avoir une meilleure connaissance de la base quand j'ai échangé dans le cadre du partenariat avec l'INSEE.

Tu es maintenant une experte de la base FIDELI !

Je te remercie Elise, bonne continuation.

Merci Karine.

*** Vincent Graciet, Marc Joubert, Nicolas Kempf (Insee), Caroline De Vellis, Elise Thouron (agence d'urbanisme Bordeaux Aquitaine).** Un ensemble de marchés de l'habitat interconnectés au sein du pôle urbain de Bordeaux. INSEE Analyses Nouvelle-Aquitaine, n°114, janvier 2022 : <https://www.insee.fr/fr/statistiques/6041795#consulter>