



HAL
open science

Rédiger un plan de gestion de données (PGD) Méthode pour structurer le recensement de données destinées à l'analyse territoriale

Mathilde Plard, Guéno le Le Berre

► To cite this version:

Mathilde Plard, Gu no le Le Berre. R diger un plan de gestion de donn es (PGD) M thode pour structurer le recensement de donn es destin es   l'analyse territoriale. 2019. hal-02132724

HAL Id: hal-02132724

<https://hal.science/hal-02132724>

Preprint submitted on 17 May 2019

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 destin e au d p t et   la diffusion de documents scientifiques de niveau recherche, publi s ou non,  manant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv s.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Rédiger un plan de gestion de données

Méthode pour structurer le recensement de données destinées à l'analyse territoriale

Mathilde PLARD,
Chercheuse CNRS
UMR ESO 6590, Université de Nantes
mathilde.plard@univ-nantes.fr

Guénoilé LE BERRE,
Étudiant M1 GAED, parcours SAGT,
stagiaire au Gérontopôle des Pays de la Loire
guenole.le-berre@etu.univ-nantes.fr



Résumé

La collecte de données structure tout travail de recherche. Cette démarche peut s'avérer longue et fastidieuse. Alors que bon nombre de données publiques sont en accès libre, il peut s'avérer compliqué de connaître et mobiliser toutes les sources disponibles. Comment justifier l'utilisation de telle donnée et plutôt qu'une autre ? L'enjeu méthodologique est de taille. Cet article propose un retour sur la mise en place d'un plan de gestion de données (le Data Management Plan) afin de donner à voir les étapes d'une collecte de données. Précisément, il s'agit d'une proposition de structuration d'un catalogue, identifiant les bases de données à mobiliser, au regard de leur contenu pour cerner les enjeux du vieillissement de la population. Concrètement, le catalogue s'accompagne d'un plan de gestion pour décrire la recherche des données. Cet article tente de mettre au clair les étapes clés auxquelles le chercheur est confronté lorsqu'il souhaite établir les soubassements du recueil de ses matériaux.

Cet article a pour vocation d'apporter un éclairage sur la démarche mise en œuvre afin de préciser les étapes clés dans la création d'un plan de gestion de données. Dans un premier temps, nous verrons quel a été le contexte d'obtention des données. Puis, nous décrivons ces données, et pour finir, nous montrerons en quoi cette méthode de collecte et de structuration des données est fiable

Mots-clefs : Plan de gestion de données ; Méthode ; Catalogue ; Pré-diagnostic - Diagnostic territorial ; DMP OPIDoR



Writing a Data Management Plan (DMP)

Method to shape the collection of data for territorial analysis

Abstract

Data collection structures all research work. This can be a long and time-consuming process. While much public data is freely available, it can be difficult to know and mobilize all available sources. How to justify the use of such and such data rather than another? The methodological challenge is significant. This article provides feedback on the implementation of one data management plan in order to show the steps involved in data collection. More specifically, it is a proposal to structure a catalogue, which identifies the databases to be mobilized, with regard to their content in order to understand the challenges of an aging population. In concrete terms, the catalogue is accompanied by a management plan to describe the data search. This article attempts to clarify the key steps that the researcher faces in establishing the foundation for the collection of his materials.

The purpose of this article is to shed light on the approach implemented in order to specify the key steps in the creation of a data management plan. First, we will see what the context was for obtaining the data. Then we will describe this data, and finally, we will show how this method of data collection and structuring is reliable.

Keywords: Data management plan; Method ; Catalogue ; Pre-diagnosis - Territorial diagnosis ; DMP OPIDoR

INTRODUCTION

Ce travail d'explicitation est réalisé dans le cadre d'un diagnostic territorial sur les enjeux du vieillissement de population à l'échelle de Mauges Communauté (Maine-et-Loire) coordonné par le Gérontopôle des Pays de la Loire¹ situé à Nantes. À l'instar du DATEL² — diagnostic action territorial environnement longévité (Chapon et al., 2012) — où la question des relations entre vieillissement et territoires est explorée en combinant une approche globale, le CAP LONGÉVITÉ met en œuvre des entrées thématiques. Pour répondre aux enjeux de faisabilité des études et assurer le bon déroulement opérationnel, cette démarche est l'occasion de cibler des leviers spécifiques selon les territoires. Ainsi, le « CAP LONGÉVITÉ Habitat » cible précisément les enjeux relatifs aux modes d'habiter des aîné.e.s.

Le diagnostic territorial est devenu un document essentiel dans la mise en œuvre de politiques publiques territoriales. Quelle que soit sa problématique (habitat, agriculture, etc.), et sans s'y limiter, il s'appuie sur des données statistiques lors de sa phase initiale, le diagnostic factuel. La réalisation de diagnostic, mais aussi de tout travail de recherche, croise des enjeux liés à la profusion des données, à l'évolution vers l'*open access* et à la nécessité de transparence dans les méthodes de collecte.

L'enjeu méthodologique est central pour assurer la rigueur de la collecte des informations et le sérieux des analyses.

La mise en place d'un annuaire de données référencées constitue un préalable nécessaire dans la mise en place du diagnostic « CAP LONGÉVITÉ Habitat ». Il propose un recensement et une description des bases de données, en fonction des variables ou indicateurs qu'elles contiennent. Face aux enjeux méthodologiques évoqués précédemment, un *plan de gestion des données* (PGD) a été réalisé. L'originalité de la méthode testée réside dans la proposition d'associer un catalogue de données et un PGD. Les objectifs recherchés grâce au catalogue sont de structurer les bases de données, selon les thématiques qu'elles traitent. Ils sont aussi de garantir la pérennité du travail. En parallèle, utiliser un plan de gestion d'après le modèle FAIR de la Commission Européenne (CE), permet de clarifier la manière dont est réalisée la collecte. C'est donc la réutilisation du travail qu'il garantit. Plus précisément, le PGD permet de retrouver les données sélectionnées (Findable). Il décrit comment y accéder (Accessibility). En outre, il vise à faire en sorte que d'autres puissent travailler et mobiliser de nouveau ces mêmes données. Il recherche donc l'Interopérabilité des données. Ainsi, il rend les données réutilisables (Re-usable).

¹ Accompagner les collectivités territoriales, dans la mise en place de projets gérontologiques sur leur territoire, est une des missions du Gérontopôle des Pays de la Loire. Basée à Nantes, il s'agit d'une association créée en 2010, à l'initiative du Professeur G. Berrut, avec le soutien du Conseil Régional et l'appui du CHU de Nantes, d'Angers et de la CCI Pays de la Loire.

² La méthodologie du DATEL été construite par des géographes de l'UMR-CNRS 6590 ESO (PIHET, PLARD), et des médecins gériatres du Gérontopôle des Pays de la Loire (BERRUT) et de l'cade (CHAPON). Le DATEL (Diagnostic Action Territorial Environnement Longévité) vise à éclairer les élus dans leur prise de décisions pour accompagner le vieillissement et préserver la qualité de vie de la population.

Contexte d'obtention des données

Concrètement, le travail vise à accompagner une démarche de diagnostic territorial. Initialement celle-ci a été nommée DATEL, pour Diagnostic Action Territorial Environnement Longévité (CHAPON et al., 2012). La méthodologie du DATEL concerne l'ensemble des thématiques touchées par le vieillissement. Elle formalise une démarche s'appuyant autant sur des données quantitatives que qualitatives. Le diagnostic factuel vise à établir un état des lieux géographiques (données statistiques, infrastructures, etc.). En parallèle, le diagnostic partagé vise à recueillir la parole d'acteurs du territoire (élus, professionnels, bénévoles). Il "intègre les éléments qualitatifs voire les contradictions qui ont pu être observées au cours du recueil d'information" (CHAPON et al., 2012, p.125). Une première application a eu lieu de 2011 à 2012, sur la commune de Bouchemaine (49). Afin de répondre de façon plus ciblée aux certaines commandes de la part d'élus locaux, le DATEL a été repensé pour pouvoir être mené avec des entrées thématiques, problématiques. Un travail de concertation entre le GÉrontopôle des Pays de la Loire, le professeur G. Berrut du CHU de Nantes et M. Plard, chercheuse CNRS a permis la rédaction en 2018 de la démarche « CAP LONGÉVITÉ ». Les données collectées, présentées dans ce travail, concernent une orientation sur l'habitat et le vieillissement — Cap Longévité Habitat pour l'EPCI³ de Mauges Communauté.

D'un point de vue pratique, le travail de collecte et de structuration d'une base de données est réalisé dans le cadre d'un stage volontaire. Il est réalisé par Guénoilé Le Berre, étudiant en première année de master, à l'Institut de Géographie de

l'Université de Nantes sous la direction de Mathilde Plard (CNRS). Le livrable restitué à l'issue du stage présente la base de données et sa notice explicative précisant les méta-données. Un travail réflexif est également mené en parallèle pour formaliser les étapes de ce procédé, le plan de gestion de données et l'écriture de cet article constituent les deux livrables de ce travail. Le stage se déroule sur le site du GÉrontopôle. Les activités sont réalisées sur ordinateur personnel. Le système d'exploitation est Windows 10 et les logiciels utilisés sont pour parti propriétaire, pour parti libre. Pour la mise en œuvre d'un PGD, il est considéré qu'un espace dédié à la gestion du projet est nécessaire (ARNOULD, JACQUEMOT, 2016, p. 7). De ce fait, aucun équivalent Freeware trouvé, Google Drive (gratuit) a été choisi, pour permettre le suivi de l'avancement du travail. Suivi effectué sur le lieu de travail au GÉrontopôle par la personne en charge de la mission formation et action territoriale. On aurait également pu penser à Basecamp, mais ce logiciel de suivi de projet est payant. Sur ce Drive a été créé un journal en format Doc et un tableur en format Sheet, pour le catalogue. Ce Drive a fait l'objet de sauvegarde régulière et sera supprimé à l'issue de l'étude. En effet, les documents produits seront stockés sur le serveur du GÉrontopôle, lui-même sauvegardé régulièrement. Les données collectées ne nécessitent pas de stockage particulier, car lorsqu'elles sont collectées, il s'agit de données publiques accessibles. Dans le cas où les données recensées et décrites nécessitent une commande ou un partenariat, il s'agit simplement d'obtenir des renseignements sur les données et sur la manière de les obtenir. En parallèle, l'utilisation d'un logiciel libre a été testée

³ EPCI : Établissement public de coopération intercommunale

pour réaliser un rétroplanning des activités (GanttProject).

Concernant le plan de gestion des données (PGD), il y avait plusieurs possibilités. En effet, plusieurs acteurs interviennent dans le monde de la recherche, pour inciter au développement de PGD dans les programmes de recherches. Leur objectif partagé est de soutenir une évolution vers l'*open access*. Les outils qu'ils développent offrent un service d'aide aux chercheurs qui souhaitent utiliser un PGD. Il semble que le premier outil à destination des chercheurs soit DMP Tool. Il a été mis en place en 2011 aux USA, par différentes institutions, telles que University of California Curation Center (UC3), Digital Curation Center (DCC-UK), Los Angeles Library et d'autres universités américaines. Il s'agissait d'une demande d'agences de financements de projets, tels que la National Science Foundation (NSF). En ligne, il est gratuit et *open source*. Au Royaume-Uni, a également été développé un outil de *plans de gestion de données*. Il s'agit de DMP Online. Cet outil de PGD semble assez facile à utiliser et adapter selon les besoins. Il est conçu pour "assister les chercheurs" avec "la création en ligne de PGD". Il s'agit d'un logiciel *open-source*, donc "customisable" (GAILLARD, 2014, p. 66). À la suite de DMP Tool et DMP Online, a été développé un nouvel outil pour garantir un caractère *open-source*, par les mêmes porteurs de projets. Il s'agit de DMP Roadmap.

La Commission Européenne de la Recherche propose un modèle de PGD appelé "FAIR" pour Findable, Accessible, Interoperable et Re-usable. Il existe d'autres types de modèles, mais celui-ci a été choisi, car il est issu de la CER, qui peut financer des projets. On peut récupérer

celui-ci sous forme de tableur Excel ou en PDF. Il nous suffit alors de remplir lignes et champs, d'en supprimer ou d'en rajouter si besoin. Il a été développé à l'initiative d'acteurs anglais et français, dans le cadre d'un programme de financements de projets de recherche par l'UE, le programme Horizon 2020 (DEBOIN, 2018, p. 4). C'est la seconde version de ce PGD FAIR qui a été utilisé. Elle a été publiée en 2018 par les trois universités Paris Diderot, Sorbonne Paris Cité, et Paris Descartes.

Le modèle FAIR a été utilisé non pas sur un tableur, mais grâce à l'outil DMP OPIDoR. Celui-ci a été développé en France, en 2016. Outil en ligne, il est disponible en français et en anglais. Son utilisation a été jugée la plus simple, au vu de l'interface, mais aussi parce qu'elle est possible en français. De plus, une fois un compte créé, on peut modifier autant que l'on veut le PGD, mais aussi le partager. Par ailleurs, il est possible de l'exporter en différents formats. Pour sa réalisation, la description des données est facilitée par des étapes, accompagnées de questions auxquelles on doit répondre. Ces questions permettent de ne rien oublier pour garantir que le PGD soit bien "FAIR".

De manière générale, il faut décrire les données dans un PGD. Pour des données existantes, ce qui a été le cas pour le *diagnostic factuel* du Cap Longévité Habitat, il faut décrire le chemin d'accès aux données. C'est-à-dire qu'il faut en donner la source, les modalités d'accès et une description synthétique (DoRANum). Dans chacun des outils de PGD évoqués précédemment, on trouve plus ou moins les mêmes rubriques à remplir, présentées dans le tableau suivant.

Tableau 1 : Contenu d'un plan de gestion de données

Rubriques	Contenu
Informations administratives	Description du projet et information permettant de le situer (financements, responsables)
Collection de données	Données utilisées et descriptions de celles-ci
Documentation et métadonnées	Informations nécessaires pour lire et interpréter les données ; organisation de la collecte ; standards adoptés
Éthique, cadre légal	Conservation, protection et partage des données
Étapes pour la sécurité du stockage et du transfert de données sensibles	Propriétaire(s) et licence(s) des données ; délai d'embargo et restrictions de partage
Stockage, sauvegarde, sécurité Stockage et sauvegarde	Lieu de stockage des données ; plan de sauvegarde ; personne(s) responsable(s) de la sauvegarde, risques et leur gestion ; dispositif d'accès
Sélection et conservation	Informations détaillées sur les données qui seront retenues, partagées et/ou conservées, et référence aux obligations contractuelles, légales ou réglementaires ; utilisations prévues ; durée de conservation des données au-delà du projet ; entrepôt
Partage des données	Étapes à mener, conditions de restriction du partage, mécanisme de partage, délai de publication
Responsabilités et moyens	Nom de la personne responsable du PGD, équipements et logiciels de gestion de données

Source : d'après DEBOIN, 2018, pp. 2-3

La recherche et la collecte de données ont été réalisées sur internet, via Google Chrome. Il s'agit soit de recherche par mots-clefs directement sur le moteur de recherche, soit de recherche sur le site de l'Insee ou d'autres fournisseurs de données. L'échelle de collecte est inférieure aux EPCI (communes, îlots regroupés pour l'information statistique - IRIS, quartier, parcelle). En effet, les diagnostics se feront à l'échelle des EPCI. Or, ceux-ci sont des constructions politiques qui s'étendent de plus en plus sur de nombreuses communes. Pour représenter au mieux la spatialisation des faits sociaux démographiques dans leur diversité, l'échelle d'analyse doit donc être inférieure.

Un postulat de la recherche de données est que des adhérents et partenaires du Gérontopôle pourraient en disposer et en partager avec celui-ci. Dans ce cas, il s'agit d'identifier la manière dont on peut les récupérer (démarche, coût, délai). L'identification d'organisme à contacter se fait par trois voies. Tout d'abord, la liste des adhérents du Gérontopôle des Pays de la Loire est consultée. Celle-ci est disponible en ligne. Puis, sont consultées les sources de publications de certains organismes sur les thématiques visées. Ces publications sont, soit des publications internes au Gérontopôle (la cartographie réalisée par

V. Guichet, de l'Atelier de l'Espace⁴, dans le cadre de la démarche Cap Longévité menée à Mauges Communauté), soit des publications externes concernant le vieillissement et l'habitat, comme celles des Agences Départementales d'Informations sur le Logement (ADIL), des Confédérations de l'Artisanat et des Petites Entreprises du Bâtiment (CAPEB), de l'Union Sociale pour l'Habitat (USH) ou de l'INSEE. Dans ces publications sont consultées les sources des données.

Pour finir, la confrontation de ces informations collectées avec le carnet d'adresses du Gérontopôle, permet d'identifier des personnes à contacter — une personne qui connaît le Gérontopôle, ses actions, et qui a potentiellement déjà travaillé avec l'association, de près ou de loin. Le postulat est alors qu'une personne sera plus encline à répondre à une demande, voire à donner un autre contact, si elle connaît son interlocuteur et ses raisons d'agir. Ces demandes de partage de données sont faites par mail. Le mail est envoyé individuellement. Il donne les mêmes informations ; il décrit le cadre du travail et ses objectifs (catalogue de bases de données pour le Gérontopôle, réalisé par un stagiaire en formation de

géographie-aménagement, dans le but de pouvoir répondre à des commandes de collectivités en Pays de la Loire); dans quelle démarche ce travail s'inscrit (le diagnostic territorial Cap Longévité); la demande (données spécifiques, conditions de partage); un contact (mail, numéro de téléphone fixe du Gérontopôle, numéro de téléphone portable personnel). Une prise de rendez-vous est proposée.

Si la réponse par mail ne donne pas de suite, une relance est prévue au plus tôt trois jours après. Si la personne redirige la demande vers une autre personne, la même démarche sera adoptée pour cette personne suivante, et ainsi de suite. Il s'agit de savoir si l'organisme a des données concernant le vieillissement (allocation, demande de logement ou d'hébergement, subventions pour rénovation du logement). C'est dans un second temps, après une première réponse, que pourront être décrits plus précisément les attendus selon les organismes contactés. La démarche de collecte des données n'est donc pas spécifiquement originale, de par le caractère des données recherchées. En effet, il s'agit de données publiques qui ont déjà été utilisées pour différentes finalités.

⁴ <https://www.latelier-de-lespace.com/>

Présentation des données

L'ensemble des données couvre des informations sur le vieillissement de la population à la commune ; l'habitat en fonction de ses occupants et des types de logements (taille, nombre de pièces ; collectif ou occupé par une personne seule) ; mais aussi sur l'ensemble de la population (évolution démographique, situation sociale et économique). L'objectif est de prendre en compte la perte d'autonomie potentielle des habitants d'un EPCI.

Les données sont des données publiques. En revanche, elles ne sont pas toujours gratuites ou accessibles en ligne. Celles accessibles en ligne sont des données de l'INSEE (Institut National de la Statistique et des Études économiques). Il existe des données payantes, car elles nécessitent des coûts de traitements. En effet, les données fiscales sont protégées par le secret statistique. Elles sont donc soumises à des délibérations de la CNIL, et à un traitement pour empêcher les reconnaissances d'individus par recoupement de données. Il s'agit des données issues des Fichiers Fonciers ou du cadastre. Il existe aussi des données payantes, car elles sont le fait d'une commande spécifique. C'est le cas des données Omphale. Il existe aussi des données que le Gérontopôle peut obtenir en passant des partenariats. Ces données sont identifiées grâce au protocole de collecte de données décrit précédemment.

La structuration du catalogue est un travail clef pour la compréhension et la réutilisation du travail. Pour rappel, celui doit permettre de savoir quelles variables ou données se trouvent dans quelles bases de données et comment y accéder.

Autrement dit, il classe les bases de données, selon les informations qu'elles contiennent et selon les thèmes qu'elles permettent de traiter. Les variables et indicateurs souhaités par le Gérontopôle sont listés en colonne. Sous chacun de ces champs, on trouve les bases de données auxquelles se référer. La description de celles-ci s'inspire du standard OAI-ORE (Open Archives Initiative Object Reuse and Exchange, <http://www.openarchives.org/ore/1.0/>).

Les informations suivantes sont donc fournies : l'URL, le format, le titre, le(s) auteur(s) et les dates de créations, modifications et publications. Pour adapter le travail aux besoins du Gérontopôle, et mis à part la thématique que permet de traiter et les indicateurs que permet d'obtenir la base de données d'autres informations sont ajoutées. Il s'agit des variables intéressantes, une courte description de la base de données (années des données, enquête ayant permis la collecte des données), l'échelle à laquelle sont disponibles les données (ex : communes), la géographie administrative à laquelle sont disponibles les données (ex : données de 2014 aux communes du premier janvier 2016), ainsi que l'accessibilité (payante/gratuite) et la description du chemin qui a permis de trouver la base de données. Cette description permet de retrouver les bases de données de différentes manières :

- par le titre
- par le chemin
- par l'URL
- par la source (diffuseur et producteur)

Toutefois, un travail de mise à jour du catalogue sera inévitable, pour actualiser les données, dès que des dates plus récentes sont disponibles.

Tableau 2 : Grille de description des données du CAP Longévité Habitat

Thématique	
Nom indicateur ou variable	
Description de l'indicateur ou variable	
variables nécessaires pour calculer ou représenter cet indicateur ou pertinentes pour décrire une situation	
Intitulé de la base de données	
Description des données collectées	
Année de publication	
Date de consultation	
Échelle	
Géographie à laquelle sont agrégées les données	
Diffuseur de la donnée	
Producteur de la donnée	
Accessibilité	
URL	
Chemin à suivre pour récupérer la donnée	
Format auquel sont disponibles les données	

La pérennité des documents produits est garantie de différente manière. En premier lieu, les documents sont enregistrés sur le serveur du Gérontopôle. Ils sont sauvegardés en format ods (pour le PGD), calc (pour le catalogue) et PDF (pour les deux). De plus, le partage de ces deux documents, en annexe de cette publication, a donné lieu au choix d'une licence pour la grille du catalogue. Il s'agit d'une licence Creative Commons. La licence est une des deux licences jugées les plus en adéquation avec la philosophie du libre. Il s'agit de la licence Creative Commons Attribution (CC BY).

Autrement dit, il s'agit d'une licence qui nécessite, lors de la réutilisation de l'œuvre produite (ici le catalogue), de citer ses auteurs. En revanche, il est possible de s'inspirer de ce travail, de le modifier, de le diffuser, y compris de l'utiliser à des fins commerciales. Ce partage ne concerne que le contenant du catalogue (la grille). Pour le contenu (les variables et indicateurs, les bases de données recensées et décrites), il reste au sein du Gérontopôle.

Fiabilité des données

Les données recensées, sur lesquelles s'appuient les diagnostics territoriaux Cap Longévité, sont des données publiques. Elles sont donc issues de grandes enquêtes (recensement de la population) ou de données administratives (données fiscales, données cadastrales). Leur collecte est donc documentée, ainsi, que la population concernée, mais aussi les limites de la collecte. Par exemple, le recensement de la population n'interroge pas l'ensemble de la population, à chaque fois. Les données MAJIC (Mise à Jour des Informations Cadastrales), ne concernent que les informations sur les propriétaires, ou les personnes physiques payant la taxe d'habitation, et non les occupants réels (Cerema Nord-Picardie, 2014, p. 16, 2017, p. 32).

De plus, recenser des données complémentaires, traitées et mises à disposition par des partenaires du Gérontopôle, peut apporter une plus-value au diagnostic. Cependant, ces données doivent être complétées par des enquêtes qualitatives. C'est effectivement ce qui est prévu dans la démarche de diagnostic partagé du DATEL (CHAPON et al., 2012) et du CAP LONGÉVITÉ.

La structuration du catalogue était l'une des difficultés de ce travail. Pour rappel, l'objectif final était de pouvoir savoir dans quelle base de données aller chercher telle variable ou tel indicateur. Le catalogue offre donc deux informations importantes. Il y a par colonne, le nom d'une variable ou

d'un indicateur et sa courte description. On trouve en dessous, la ou les bases de données dans lesquelles on peut trouver l'information recherchée. Puis, chaque base de données est décrite (courte description, dates, échelles, sources, accessibilité dont format). Ces informations permettent de retrouver les données. Cependant, il est possible d'émettre l'hypothèse que l'utilisateur puisse rencontrer des difficultés de lecture. La taille du catalogue tout d'abord, dans un format tableur, ne permet pas de visualiser, en une seule fois, l'ensemble des informations renseignant un champ. De plus, les courtes descriptions de l'indicateur et de la base de données peuvent ne pas être adaptées à n'importe quel utilisateur. En effet, il est supposé qu'il faille une connaissance minimum des variables, de leur utilisation possible, mais aussi des données. En outre, rechercher des données pour compléter le catalogue s'est révélé chronophage. Ce qui souligne l'intérêt de réaliser un document pérenne.

Le document doit donc pouvoir rester adaptable (pouvoir ajouter des champs, des bases de données), ce qui exclut par exemple un format PDF pour le stockage. Dans ce sens, le travail aurait pu être mené de manière différente. Il aurait par exemple été possible de travailler avec d'autres outils, dans une autre philosophie de logiciel. Il aurait donc été possible d'utiliser le système d'exploitation Linux, pour pouvoir construire le catalogue avec le logiciel *Freeware* de tableur, GNUméric.

CONCLUSION

Les enjeux méthodologiques sont toujours importants, lorsqu'il s'agit de choisir des données, mais aussi de structurer leur stockage. Dans le cas présent, il s'agissait de recenser des bases de données selon les informations qu'elles contiennent, et de les décrire selon plusieurs principes. Il fallait fournir une information minimum pour donner des connaissances à l'utilisateur sur la donnée, mais également dire comment y accéder. Face à ces enjeux, l'occasion s'est présentée de chercher à structurer un recensement de bases de données pour permettre leur usage dans des travaux futurs. Cet article avait alors pour vocation de présenter une manière de faire. L'objectif était d'apporter une aide à celles et ceux qui souhaiteront structurer leur recherche en s'inspirant d'une démarche déjà réalisée.

Un catalogue en format tableur a été réalisé pour recenser et décrire les données. En amont et en parallèle, un plan de gestion de données a été rédigé. Il a été construit d'après le modèle FAIR Horizon 20/20 de la CER, sur l'outil en ligne DMP OPIDoR. La description du contexte d'obtention des données permet au lecteur d'avoir un aperçu de l'adéquation du travail réalisé avec son propre travail, du point de vue de la méthodologie de collecte. La description des données lui permet aussi de savoir si ce travail s'appliquerait à ses données. Pour finir, la fiabilité du travail est discutée.

Il s'agit donc d'une proposition de méthodologie, qui prend du temps et nécessite sans doute des adaptations. Malgré tout, elle permet de chercher à répondre du mieux possible, aux enjeux de profusion des données à collecter, de la transparence de cette collecte et de partage du travail effectué. Le travail offre d'ailleurs des possibilités d'améliorations identifiées. En effet, il sera nécessaire d'actualiser les bases de données, lorsqu'elles sont modifiées par leur auteur (ajout de champ, changement de méthode de collecte, publication plus récente). Ainsi, consulter la page web de l'INSEE sur la publication de leurs données, permettra de savoir quand actualiser ces données. Il sera également possible le travail à d'autres problématiques que celle de l'habitat (mobilité, services à la personne, alimentation). À ce propos, certaines informations recensées dans le catalogue sont communes à plusieurs thématiques (données sociodémographiques). En outre, il offre la possibilité au Gérontopôle de signer de nouveaux partenariats, pour répondre à de prochaine demande d'EPCI. À terme, si le catalogue se révèle opérationnel dans les prochaines utilisations, le partager sur le site du Gérontopôle peut être envisageable. La meilleure manière de savoir s'il est opérationnel sera de l'utiliser pour la mise en œuvre des prochains diagnostics Cap Longévité.

Bibliographie

ARNOULD, P.-Y., & JACQUEMOT, M.-C. (2016). Guide de bonnes pratiques Gestion et valorisation des données de la recherche. OTELo ; INIST-CNRS. Retrieved June 8, 2018, from <https://hal.archives-ouvertes.fr/hal-01275841>

Cerema Nord-Picardie, 2014. "Recensement des usages et études locales s'appuyant sur les Fichiers fonciers". [en ligne]. From https://datafoncier.cerema.fr/system/files/documents/2018/06/EtudesetudesCeremaVdef_cle7bfd6f.pdf

Cerema Nord-Picardie, 2017. "Guide de prise en main des Fichiers fonciers". [en ligne]. From <http://www.geoinformations.developpement-durable.gouv.fr/guide-de-prise-en-main-des-fichiers-fonciers-et-a2156.html>

CHAPON, P.-M., PIHET, C., JAHAN, F., MICHEL, B., RIOBE, A.-L., MERJAGNAN-VILCOCCQ, C., PLARD, M., BERRUT, G., 2012. "Le diagnostic action territorial environnement longévité (Datel) : un diagnostic territorial pour construire un avenir commun avec les aînés". *Geriatr Psychol Neuropsychiatr Vieil*, 10(2), pp. 123-127

DoRANum. (2017a). Data Management Plan (DMP), Fiche Synthétique. doranum.fr. Retrieved June 6, 2018, from <http://doranum.fr/fiche-synthetique-plan-gestion-donnees-dmp/>

DoRANum. (2017b). La minute plan de gestion de données (Data Management Plan). Doranum BSN. Retrieved June 6, 2018, from <https://www.youtube.com/watch?v=5RNYGGrotZo&t=2s>

GAILLARD, R. (2014). De l'Open data à l'Open research data : quelle(s) politique(s) pour les données de recherche ? Retrieved June 6, 2018, from <http://eprints.rclis.org/22746/>

Open Archives Initiative Object Reuse and Exchange. (14.08.2014). ORE Specifications and User Guides. Table of Contents. Retrieved June, 2018, From <http://www.openarchives.org/ore/1.0/>

IST - Groupe PGD. (2015). Datapartage - Plan de gestion. INRA. Retrieved June 5, 2018, from <https://www6.inra.fr/datapartage/Generer/Plan-de-gestion/>