



HAL
open science

Pour un observatoire des données géographiques du Web. Expérimentation à partir des infrastructures de données géographiques françaises

Matthieu Noucher, Françoise Gourmelon, Claramunt Christophe

► **To cite this version:**

Matthieu Noucher, Françoise Gourmelon, Claramunt Christophe. Pour un observatoire des données géographiques du Web. Expérimentation à partir des infrastructures de données géographiques françaises. *Revue Internationale de Géomatique*, 2019, 10.3166/rig.2019.00074 . halshs-02431590

HAL Id: halshs-02431590

<https://shs.hal.science/halshs-02431590>

Submitted on 8 Jan 2020

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.

[VERSION AUTEUR AVANT MISE EN PAGE DE L'EDITEUR]

Noucher Matthieu, Gourmelon Françoise, Claramunt Christophe (2019). « Pour un observatoire des données géographiques du Web. Expérimentation à partir des infrastructures de données géographiques françaises », **Revue Internationale de Géomatique**, Volume 29, Numéro 1, janvier-mars 2019, pp. 9-30.
<https://doi.org/10.3166/riq.2019.00074>

Pour un observatoire des données géographiques du Web.

Expérimentation à partir des infrastructures de données géographiques françaises.

Matthieu Noucher¹, Françoise Gourmelon², Christophe Claramunt³

1. CNRS - Laboratoire PASSAGES (UMR 5319) - Bordeaux

Maison des Suds, 12 esplanade des Antilles, 33607 Pessac, France

2. CNRS, Univ Brest - Laboratoire LETG (UMR 6554)

IUEM, rue Dumont d'Urville, 29280 Plouzané, France

3. Institut de Recherche de l'Ecole navale (EA 3634), Lanvéoc, 29160, France

Auteur correspondant : matthieu.noucher@cnrs.fr

RESUME. L'extension des usages de l'information géographique génère une profusion de données qui circulent sur de multiples plateformes Web. Accompagner le suivi du contenu de ces plateformes relève aujourd'hui de la gageure tant elles s'apparentent à des boîtes noires informationnelles et algorithmiques soumises à de fortes dynamiques. Le concept d'observatoire de l'information géographique semble alors approprié pour suivre, analyser et comprendre l'évolution du volume et des flux de données géographiques disponibles et leur utilisation sur une portion de territoire. Néanmoins, sa mise en œuvre soulève des questions méthodologiques que nous discutons à partir de l'expérimentation d'une observation diachronique du patrimoine de données issues des Infrastructures de Données Géographiques françaises. Nous soulignons en particulier l'importance d'un suivi régulier de ces dispositifs.

ABSTRACT. Extensive uses of geographic information generate a data deluge on many Web infrastructures. Monitoring and analyzing the content of these platforms is still a research challenge because these can be considered as very dynamic informational and algorithmic black boxes. The concept of "geographical information observatory" appears as an appropriate mean to monitor, analyze and then provide a better understanding of the evolution of the volume and flows of the spatial data available on a specific web infrastructure. Nevertheless, its implementation raises several methodological questions discussed in this paper, starting from the experimentation of a diachronic data observation of French Spatial Data Infrastructures. Finally, the importance of regular monitoring of these systems is also highlighted.

MOTS-CLES : Infrastructure de Données Géographiques (IDG) ; Observatoire de l'Information Géographique ; Analyse réseau ; Analyse spatiale ; Analyse diachronique.

KEYWORDS : Spatial Data infrastructure (SDI) ; Geographic Information Observatory ; Network analysis ; Spatial analysis ; Diachronic analysis.

1. Introduction

Au-delà des considérations sur le *data deluge* ou la « révolution géospatiale »¹, l'un des enjeux de recherche actuel en géomatique est d'être en capacité de dresser un panorama des données géographiques pour mieux comprendre à la fois les opportunités offertes par des sources devenues multiples, mais aussi les lacunes informationnelles qui existent sur certains territoires ou certaines thématiques.

Fort de ce constat, deux workshops ont été organisés lors de la 8^e conférence GIScience à Vienne (Autriche) en 2014 et de la conférence COSIT à Santa Fe (USA) en 2015, autour des *Geographic Information Observatories*. Tel que discuté lors de ces événements, ce concept inédit répondait à un objectif assez large dans la mesure où il abordait les informations géographiques et temporelles en tant que telles, mais aussi leurs utilisateurs et leurs motivations, ainsi que l'outillage théorique et méthodologique sous-jacents (Adams and Gahegan, 2014 ; McKenzie et al., 2015 ; Miller, 2017). La nécessité d'explorer *l'univers de l'information géographique* en déployant des protocoles d'observation qui permettent un suivi dans le temps des données en ligne, en termes de contenus et d'usages a également été soulignée par Ballatore (2014). Alors que l'information géographique est régulièrement mobilisée dans les phases de collecte, de traitement et de restitution des indicateurs issus des observatoires territoriaux (de Sède-Marceau et al., 2018), le développement du géoweb permet aujourd'hui d'envisager la mise en place d'observatoires de l'information géographique pour évaluer sur un territoire et/ou une thématique spécifique l'état des lieux des données géographiques disponibles. À l'instar des observatoires scientifiques qui ont pour objectif de suivre, analyser et comprendre, sur une portion d'espace représentative d'une entité territoriale, les dynamiques socio-environnementales résultant d'interactions entre systèmes socio-économiques et biophysiques (Libourel et al., 2009), nous définissons les observatoires des données géographiques comme des dispositifs sociotechniques en réseau. Ils sont mis en œuvre pour d'une part suivre, analyser et comprendre l'évolution spatio-temporelle des données géographiques sur un territoire afin de révéler les représentations spatiales associées à ses dynamiques socio-environnementales, et d'autre part, pour observer les théories, méthodologies et approches mises en œuvre dans l'utilisation de ces données géographiques. Lieu de production, d'échange et de partage d'informations et de connaissances qui s'inscrit dans la durée (Janowicz et al., 2014), ces observatoires s'organisent sur la base d'un contenu numérique et de ses interactions avec les utilisateurs, et auraient pour principales fonctions de rendre visible et de mémoriser un patrimoine immatériel difficile à saisir et souvent évanescent car malléable.

De notre point de vue, la mise en œuvre d'observatoires des données géographiques soulève au moins quatre questions méthodologiques préalables : 1) comment, face à des contenus qui circulent sur différentes plateformes web être en mesure d'identifier les données géographiques accessibles ? 2) une fois identifiées, comment les qualifier ? 3) comment en assurer le suivi dans le temps pour les voir

¹ Nous reprenons ici l'intitulé d'un web-documentaire de l'Université Penn State (USA) dont le titre et le contenu illustrent bien le discours (caricatural) de rupture qui accompagne la promotion de ces développements technologiques : <http://geospatialrevolution.psu.edu/>

évoluer, et enfin 4) quels protocoles d'observation construire pour identifier les pratiques, méthodes et outils d'analyses mis en œuvre. La composante temporelle nous semble particulièrement importante puisque l'ambition de déploiement d'un observatoire des données géographiques ne peut se concevoir comme une simple photographie à un instant « t » mais bien comme un dispositif d'accompagnement – c'est-à-dire en lien direct avec les observatoires et plateformes, notamment institutionnelles, existantes – qui s'inscrit dans la durée. Il s'agit également non seulement de mettre en œuvre des outils d'observation des données, mais aussi d'évaluer sur la durée ce que les utilisateurs en font.

Dans cette contribution, nous proposons d'expérimenter un protocole d'identification, de qualification et de suivi de l'évolution des données issues des Infrastructures de Données Géographiques (IDG) françaises. Les IDG sont, en effet, devenues des plateformes web essentielles pour la diffusion des données institutionnelles (Rajabifard et al., 2003). Elles visent à faciliter, à tous les échelons territoriaux, la découverte, l'accès, l'échange et le partage d'information et de services géographiques (Masser, 2010). Les IDG contribuent à la complexité de l'*univers de l'information géographique* (Ballatore, 2014 ; Janowicz et al., 2014) et sont considérées par certains auteurs (Grus et al., 2010) comme des systèmes adaptatifs qui ajustent leur contenu, leurs actions et objectifs en fonction de facteurs externes.

Notre corpus est constitué de 39 IDG nationales et régionales qui disposent d'un géocatalogue opérationnel. A travers leur exploration, nous produisons des indicateurs concernant leur couverture spatiale, thématique ou encore organisationnelle à deux dates (2016 et 2017). Ce faisant, cette expérimentation peut s'apparenter à une tentative de prototypage d'un observatoire de l'information géographique issue des IDG dont nous soulignons, en conclusion, les enjeux et limites.

2. Méthodologie

2.1. Extraction des données

Pour extraire les informations contenues dans les métadonnées des 39 IDG étudiées², un script Python³ 2.7 utilisant la bibliothèque OWSLib⁴ a été développé (Rouan et Pierson, 2016). Cette bibliothèque offre la possibilité d'interroger des catalogues en utilisant les différents standards de l'Open Geospatial Consortium (OGC), dont le service web de catalogage (CSW) dans sa version 2.0.2. Pour cette phase d'extraction, plusieurs problèmes d'ordre technique ont été rencontrés. Selon la plateforme utilisée par l'IDG (GeoNetwork, Isogeo, Prodige, Amigo...), l'implémentation du CSW peut varier ; une IDG pouvant ne plus être interrogeable

² L'identification de ces IDG régionales et nationales a été réalisée à partir de l'annuaire de l'AFIGEO en vérifiant l'opérationnalité des plateformes. La liste complète est disponible en annexe.

³ <https://github.com/LETG/csw-harvester>

⁴ <https://github.com/geopython/OWSLib>

(ré-indexation, problème réseau) pendant l'exécution du script. Il faut donc pouvoir relancer régulièrement la procédure. Les CSW de notre corpus ont donc été testés sur une période de plusieurs semaines à chaque fois : en juillet pour l'extraction de 2016 et en mai celle de 2017. D'autres problèmes inhérents aux contenus des fiches de métadonnées ont été identifiés. Par exemple toutes les rubriques, même celles rendues obligatoires par la norme ISO 19115, ne sont pas toujours remplies ou peuvent être mal renseignées. C'est le cas des dates (création, publication, révision) ou encore de la généalogie dont le champ de saisie libre s'est avéré particulièrement mal renseigné pour permettre une analyse systématique.

La librairie OWSLib a permis d'importer les fichiers XML des métadonnées et de réaliser leur analyse avec des requêtes XQuery. Cependant le souhait de rendre ces données plus accessibles notamment en proposant des analyses et des synthèses dynamiques à travers le Web nous a conduit à les structurer dans une base de données relationnelle. Cette phase d'archivage est réalisée au moyen de la bibliothèque Psycopg⁵, adaptateur PostgreSQL le plus utilisé en Python.

2.2. Archivage et gestion temporelle des données

Concernant l'archivage des données, l'objectif n'est pas d'explorer l'ensemble des métadonnées mais de les analyser au regard de plusieurs questions de recherche (tableau 1). Celles-ci sont structurées autour des quatre objectifs prioritaires mis en avant par les promoteurs des Infrastructures de Données Géographiques françaises, que nous avons interrogés⁶. Ces derniers mettent souvent en exergue la capacité de leurs dispositifs à mettre en réseau les Systèmes d'Information Géographique existants. En brisant les silos grâce notamment à l'interopérabilité des systèmes et à la normalisation des données, les IDG diminueraient les risques de rétention d'information, limiteraient les jeux de pouvoir autour des données et faciliteraient les dynamiques collaboratives. Les promoteurs des IDG s'inscrivent ainsi pleinement dans l'idéologie du « libéralisme informationnel » théorisée par Benjamin Loveluck (2015) à partir de son travail de généalogie politique d'Internet. En effet, ils légitiment le déploiement des IDG pour améliorer la circulation des données, cette dernière étant érigée en vecteur de démocratie (plus de données serait synonyme de plus de transparence) et de développement économique (plus de données permettrait de générer plus d'innovation).

⁵ <http://initd.org/psycopg/>

⁶ 12 entretiens semi-directifs menés au début du projet (2016) auprès de responsables d'IDG nationales (6) et régionales (6).

Tableau 1. Les questions de recherche associées aux quatre objectifs mis en avant par les promoteurs des Infrastructures de Données Géographiques en France.

Accessibilité des données	<ul style="list-style-type: none"> a. Quels sont les discours des promoteurs d'IDG sur l'accessibilité des données ? b. Quelles sont les données géographiques effectivement diffusées ? c. Les données géographiques diffusées sont-elles réutilisables ?
Interopérabilité des outils	<ul style="list-style-type: none"> a. Quels sont les discours des promoteurs d'IDG sur l'interopérabilité des outils ? b. Les interconnexions entre infrastructures permettent-elles d'accélérer les flux de données ? c. Que change l'interopérabilité pour les usagers ?
Mise en réseau des acteurs	<ul style="list-style-type: none"> d. Quels sont les discours des promoteurs d'IDG sur la collaboration entre acteurs ? e. Les IDG favorisent-elles la coproduction de données géographiques ? f. Les IDG favorisent-elles l'émergence des communautés de pratique ?
Equité informationnelle des territoires	<ul style="list-style-type: none"> g. Quels sont les discours des promoteurs d'IDG sur l'équité informationnelle des territoires ? h. Les IDG favorisent-elles une couverture informationnelle homogène de leur territoire ? i. Que change l'homogénéité/hétérogénéité de la couverture des données pour les usagers ?

Pour répondre à ces questions par l'exploration des métadonnées associées au géocatalogue des IDG étudiées, nous avons donc constitué une base de données. Plutôt que de reprendre l'ensemble de la norme de métadonnées, une sélection d'éléments pertinents a été réalisée et un modèle entité-association a été élaboré. Celui-ci intègre une dimension temporelle en relevant une date d'extraction pour chaque opération (figure 1).

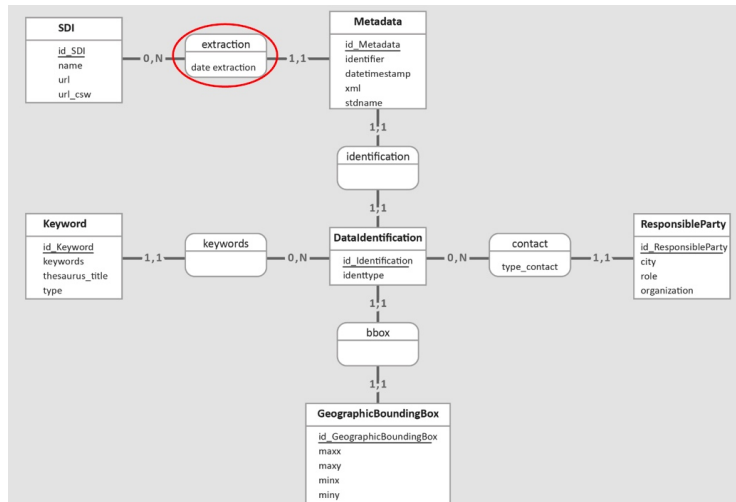


Figure 1. Modèle conceptuel de données (MCD) ayant servi à implémenter la base d'analyse des métadonnées. La classe Extraction permet de disposer d'une information temporelle indispensable pour une analyse diachronique.

Afin de rendre la base de données évolutive, nous avons fait le choix de garder la structuration et le nom des tables/champs de la norme. Ainsi, le MCD est organisé autour de 3 entités clés (« Metadata », « DataIdentification » et « GeographicBoundingBox »). Le choix a également été fait de conserver l'ensemble des métadonnées en XML dans un champ de la table « Metadata » permettant ainsi de ne perdre aucune information qui pourrait être utile pour re-peupler la base en cas d'évolution. La base de données a été implémentée sous PostgreSQL 9.3.

2.3. Choix des indicateurs de suivi

Le prototype d'un observatoire des géocatalogues des IDG françaises que nous proposons repose sur des indicateurs qui visent justement à interroger les apports effectifs des IDG en termes de mise en réseau des données, des outils, des acteurs, des territoires. Pour dépasser les éléments de langage des acteurs, plusieurs méthodes d'analyse (spatiale, statistique, réseau) ont été mobilisées afin de produire des indicateurs sur le contenu même de ces infrastructures. Elles seront présentées ultérieurement en association avec chaque type de résultat.

A titre illustratif nous en proposons quatre (tableau 2) qui ont fait l'objet d'une restitution sous la forme de (géo)visualisations pour l'année 2016 (Noucher et al., 2017). Cependant, comme évoqué en introduction, l'un des principaux enjeux des observatoires de l'information géographique est de dépasser la situation à un instant « t » pour suivre le patrimoine de données dans le temps et évaluer son évolution. Nous avons donc veillé aux possibilités de reproduction de l'analyse à une date ultérieure.

Tableau 2. Exemples de fiches méthodologiques pour 4 indicateurs.

	Mise en réseau des DONNEES	Mise en réseau des OUTILS	Mise en réseau des ACTEURS	Mise en réseau des TERRITOIRES
Présentation générale				
Titre indicateur	Accessibilité des données	Interopérabilité des catalogues	Couverture organisat ^o des données	Couverture spatiale des données
Définition	Part des données téléchargeables et réutilisables sans condition.	Interconnexion des géocatalogues des IDG et des autres plateformes Web.	Part des différents types de contributeurs des géocatalogues	Distribution géographique de l'emprise des données diffusées
Objectifs	<i>Visualiser et suivre l'évolution</i> de la proportion de données effectivement utilisables pour dépasser le comptage des métadonnées.	<i>Visualiser et suivre l'évolution</i> du réseau de moissonnage pour comprendre comment circule la donnée.	<i>Visualiser et suivre l'évolution</i> des types d'organisation (autorités publiques, sociétés privées...) qui alimentent les géocatalogues.	<i>Visualiser et suivre l'évolution</i> de la densité de données disponibles sur les territoires (hots spots informat ^o , zone blanches...).
Données en entrée	Métadonnées des géocatalogues.	Liens moissonneurs / moissonnés	Métadonnées des géocatalogues.	Métadonnées des géocatalogues.
Sources	Service web CSW	Enquête en ligne	Services web CSW	Services web CSW
Modalité de calcul				
Traitement statistiques	Comptage et pourcentage	Analyse réseaux	Comptage et pourcentage	/
Traitement géométriques	/	/	/	Analyse spatiale
Traitement sémantiques	Filtre sur tags <i>opendata / donnée ouverte</i>	/	Catégorisation dans une typologie <i>ad hoc</i>	/
Processus d'actualisation				
Fréquence	Annuelle	Bi-annuelle	Annuelle	Annuelle
Méthode	Relance du script	Enquête en ligne	Relance du script	Relance du script
Source	Service web CSW	Promoteur IDG	Service web CSW	Service web CSW
Interprétation des résultats				
Comparaison des résultats	Analyse diachronique Comparaison volume de métadonnées VS volume de données dites ouvertes	Analyse diachronique	Analyse diachronique	Analyse diachronique
Echelle d'analyse	Analyse globale, par échelon (régional, national), par IDG	Analyse globale sur l'ensemble du réseau des IDG	Analyse globale, par échelon (régional, national), par IDG	Analyse globale, par échelon (régional, national), par IDG
Pondération des résultats	A pondérer avec la part des URL opérationnels	/	/	/
Combinaison possible	Avec l'ensemble des autres indicateurs	/	Avec l'ensemble des autres indicateurs	Avec l'ensemble des autres indicateurs
Représentation				
Tableau	Comptage Pourcentage	/	Comptage Pourcentage	/
Graphiques	Courbe d'évolution	/	Diagramme circulaire	/
Cartes	Distribution de la densité des données	/	/	Cartes des hots spots informationnels
Autres	/	Graphe	/	/

3. Résultats

Naviguer sur les sites web des Infrastructures de Données Géographiques ne permet pas d'identifier de manière exhaustive le patrimoine de données disponible en ligne. Les interfaces de navigation des géoportails n'offrent souvent que des possibilités fonctionnelles limitées pour visualiser un nombre restreint de couches. Par ailleurs, l'entrée sur les géocatalogues par des moteurs de recherche ne donne accès qu'à un ensemble filtré de métadonnées dépendant des paramètres des requêtes. Présentés sous la forme de liste, dont l'ordonnancement est d'ailleurs rarement explicite, les résultats obtenus *via* ces moteurs de recherche n'offrent aucune possibilité de post-traitement pour explorer ces corpus. Ainsi le BRGM et Etalab, la mission gouvernementale pour le développement de l'open data, proposent des tableaux de bord (*dashboard*) depuis 2017. Ils s'inscrivent dans une logique d'évaluation et sont centrés sur des cibles restreintes : uniquement les données géographiques qui sont dans le périmètre de la Directive INSPIRE pour le dashboard du BRGM⁷, uniquement les données en open data pour celui d'Etalab⁸. Par ailleurs, en fournissant un panorama ponctuel des métadonnées, ces tableaux de bord n'assurent pas l'archivage et l'analyse des indicateurs produits. Si certains indicateurs sont répétés avec des fréquences régulières (tous les 3 ans dans le cadre du rapportage de la Directive INSPIRE, par exemple), les méthodes semblent loin d'être stabilisées :

« Cette évolution peut être relativisée par une modification de la méthodologie appliquée cette année dans le calcul de cet indicateur. Jusqu'à 2015, seul le premier thème renseigné dans les métadonnées était pris en compte pour calculer le nombre de métadonnées de données. En 2016, les séries de données sont comptées autant de fois qu'elles référencent dans leurs métadonnées de thèmes INSPIRE, augmentant ainsi artificiellement le nombre de métadonnées de données. A titre de comparaison, le nombre de métadonnées de données passe de 25501 (selon l'ancienne méthode) à 30452 (selon la nouvelle méthode). »

A propos de l'indicateur MD11, p.4-5 du « Rapport de la France sur la mise en œuvre de la directive INSPIRE pour la période 2013-2015 (réf. 2013-2015) »

Pour dépasser ces limites, la répétition selon les mêmes protocoles des opérations d'extraction et d'archivage (2.1 et 2.2) nous a conduit à appliquer les mêmes méthodes de qualification (2.3) des métadonnées, permettant *de facto* une analyse diachronique de nos quatre indicateurs de mise en réseau.

3.1. Mise en réseau des données : évolution de l'accessibilité des contenus

L'analyse des données accessibles (téléchargeables et réutilisables sans contrainte) au sein des IDG permet de nuancer le discours sur le « déluge de données ». En effet si l'indexation de métadonnées connaît une croissance exponentielle, les données réellement accessibles restent minoritaires (tableau 3). L'analyse de l'évolution 2016-2017 prend ici tout son sens puisqu'elle met en évidence des soldes négatifs qui

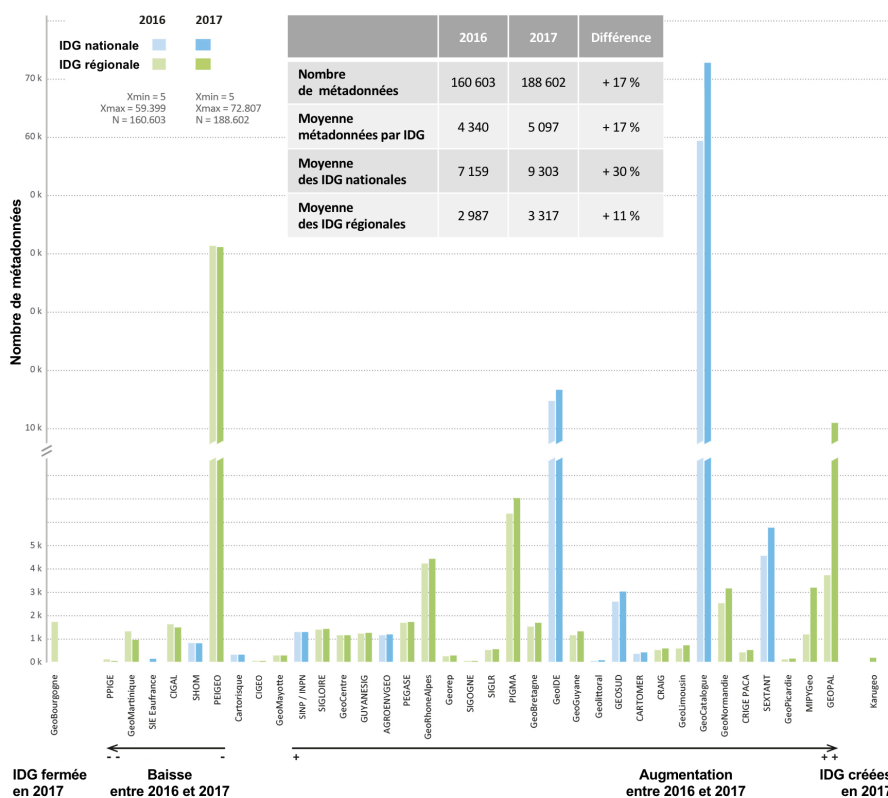
⁷ <http://dashboard.geocatalogue.fr/daobs/dashboard/>

⁸ <https://www.data.gouv.fr/fr/dashboard/>

s'expliquent par le fait que la fusion des régions⁹ a conduit à des reconfigurations ou au report de données sur des portails *open data* (figures 2 et 3).

Tableau 3. Évolution des volumes de métadonnées et de la part des données accessibles dans les IDG entre 2016 et 2017.

	Jun 2016	Mai 2017	Différentiel
Volume total de métadonnées	160603	188602	+17 %
Moyenne de métadonnées par IDG	4340	5097	+17 %
Moyenne des métadonnées dans les IDG nationales	7159	9303	+30 %
Moyenne des métadonnées dans les IDG régionales	2987	3317	+11%
Part des « données ouvertes » / « open data »	15,7 %	19,1%	+3,4 pts
Part moyenne des données ouvertes dans les IDG nationales	8,5%	22,7%	+13,2 pts
Part moyenne des données ouvertes dans les IDG régionales	19,1%	10,5%	- 8,6 pts



Source : Service web (CSW-2) des catalogues des IDG, été 2016 / été 2017

GEOS - CNRS 2017.

Figure 2. Nombre de métadonnées dans les 39 IDG étudiées en 2016 et 2017.

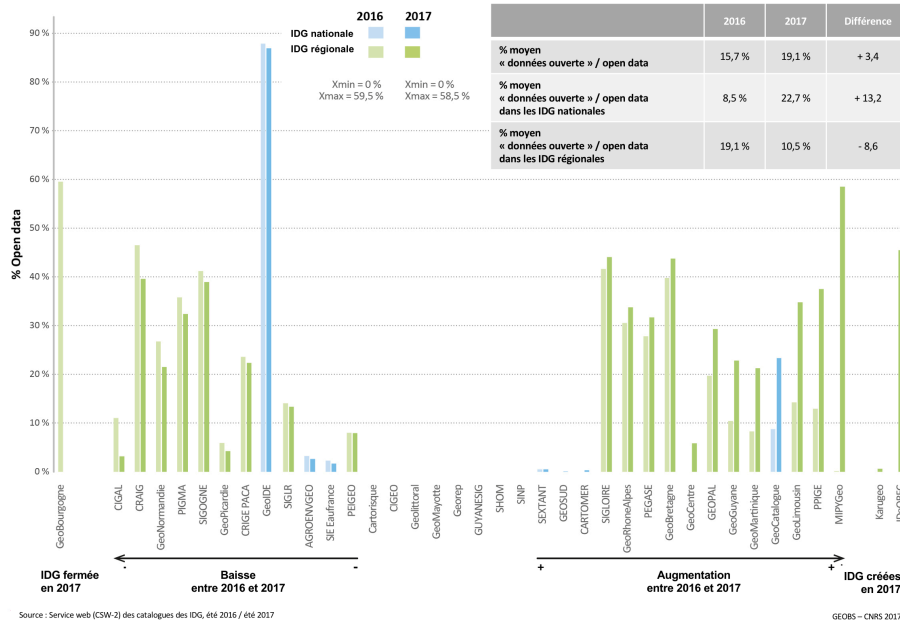


Figure 3. Nombre de données ouvertes dans les 39 IDG étudiées (2016-17).

3.2. Mise en réseau des outils : évolution de l’interopérabilité des géocatalogues

Pour évaluer l’interconnexion des IDG françaises, nous réalisons une analyse réseau du moissonnage de leurs géocatalogues. Les graphes qui en résultent pour 2015 et 2017 (figure 4 et 5) permettent d’identifier les nœuds centraux (ou isolés) et d’éventuels *clusters*. Des composantes géographiques similaires aux deux années sont mises en évidence à travers la relation des IDG régionales aux plateformes infrarégionales. En l’occurrence, les jeux d’échelle sont très marqués à travers la structure centre-périphérie du graphe. La plupart des IDG nationales occupe une place centrale. Une première couronne périphérique donne un panorama des IDG régionales qui renvoient elles-mêmes vers une deuxième couronne périphérique composée majoritairement d’IDG initiées par des collectivités territoriales, des associations ou des syndicats départementaux ou intercommunaux. De même, des composantes thématiques du réseau apparaissent à la fois en 2015 et en 2017 à travers les mises en relation d’IDG travaillant dans le domaine de la mer ou du côtier et sur des territoires ultramarins. Enfin, certaines infrastructures utilisant peu le moissonnage (nœuds de petite taille) occupent pourtant, depuis 2015, une place centrale car elles sont très moissonnées. C’est le cas d’EauFrance et de Géo-IDE.

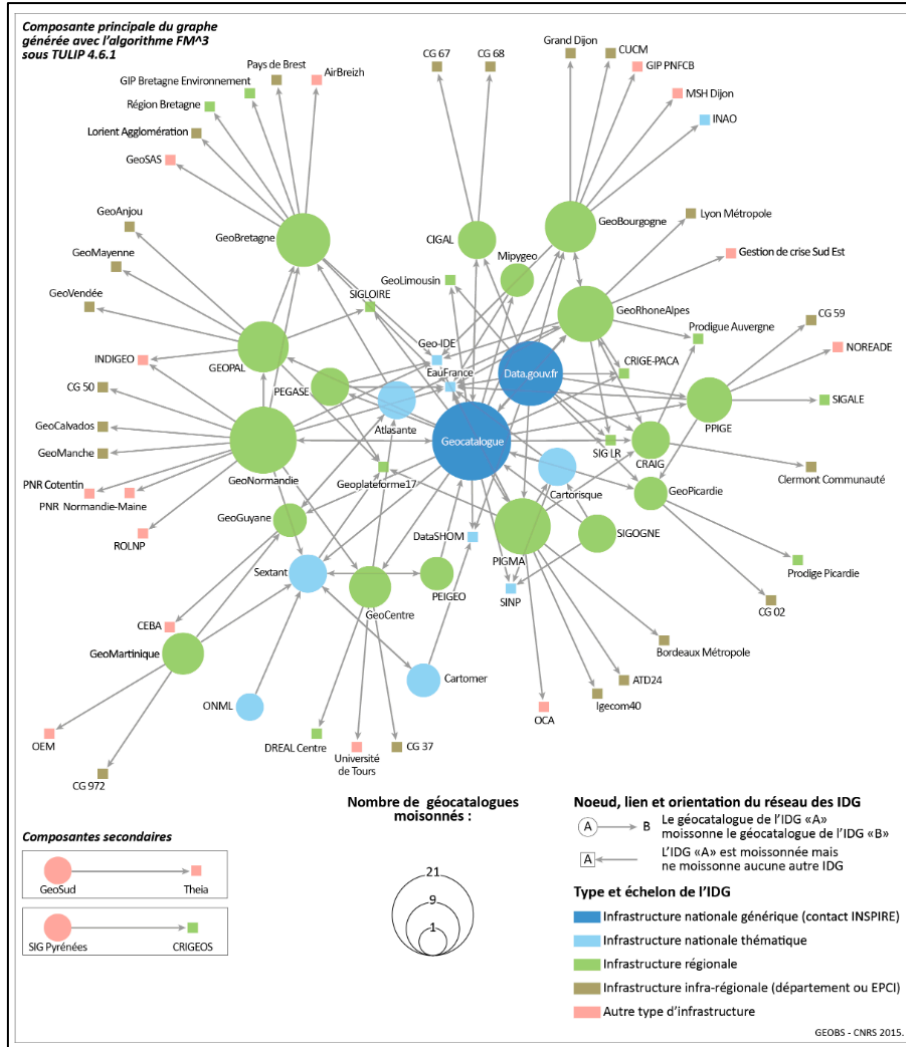


Figure 4. Réseau de moissonnage en 2015.

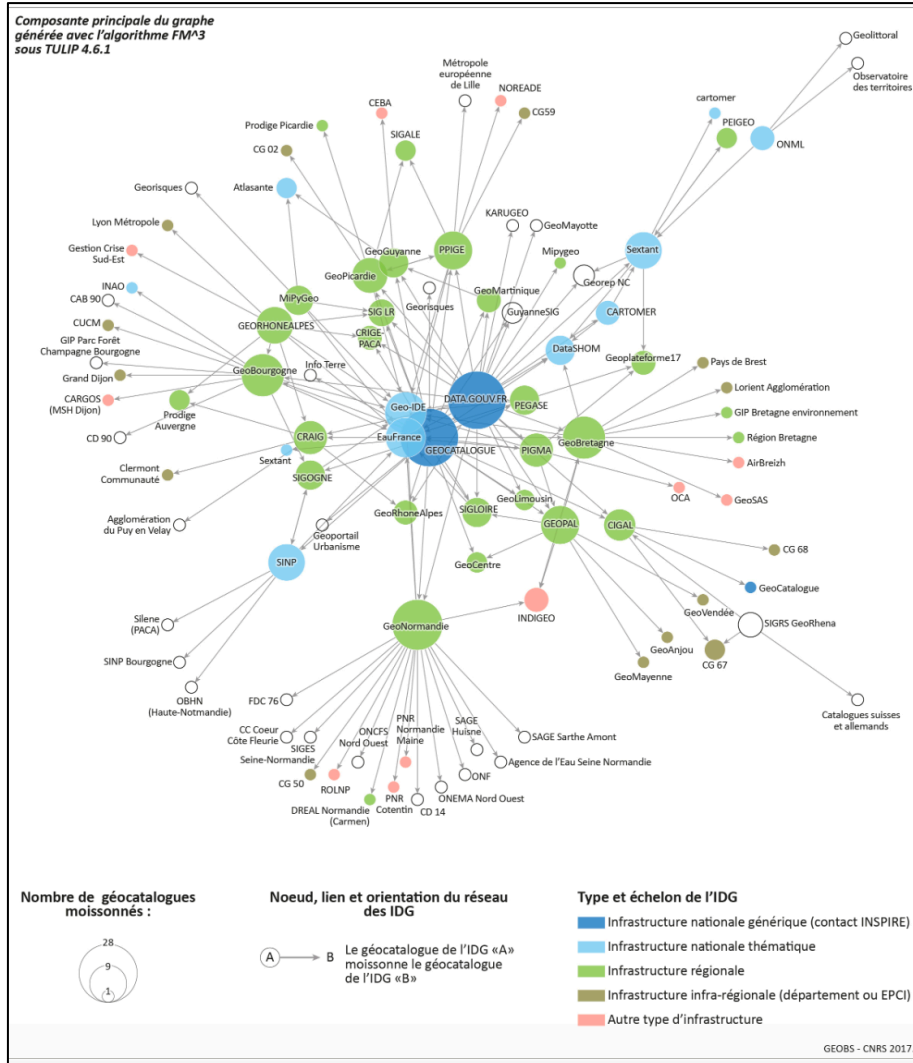


Figure 5. Réseau de moissonnage en 2017.

Si les réseaux de moissonnage de 2015 et 2017 partagent des caractéristiques communes¹⁰, l'analyse diachronique permet de pointer quelques évolutions. On peut ainsi noter en 2017 l'apparition dans le réseau d'IDG étrangères par l'intermédiaire de géocatalogues transfrontaliers ainsi que l'extension du moissonnage vers des systèmes infra-départementaux (intercommunalités essentiellement). La distribution de l'influence dans le réseau a évolué. L'indicateur moyen de *centralité*

¹⁰ Par exemple, la *densité d* qui mesure la complétude d'un graphe ($d=1$ si tous les liens possibles existent) est quasi similaire : $d_{2016}=0,022$ et $d_{2017}=0,02$.

*d'intermédiation*¹¹ a décliné entre 2015 et 2017 en raison d'une forte baisse de la valeur de cet indicateur pour le Géocatalogue national. Si celui-ci apparaissait comme le principal *hub* (nœud ayant le plus fort degré entrant) en 2015 (21 moissons), il est dépassé en 2017 par le portail data.gouv.fr (28 moissons). Cette évolution traduit bien les reconfigurations en cours entre IDG et portails *open data*.

3.3. Mise en réseau des acteurs : évolution de la couverture organisationnelle

L'analyse des métadonnées révèle également la très faible diversité des contributeurs des IDG (figure 6). Sur les 230000 contacts recensés dans les métadonnées de 2016, 70% proviennent des autorités publiques et seulement 3% du secteur privé. Cette proportion a faiblement évolué en 2017 malgré les discours volontaristes sur l'ouverture de ces infrastructures professionnelles vers l'information géographique volontaire (VGI). En revanche, la part des données issues de l'enseignement et de la recherche a fortement augmenté (de 11,7 à 15,9%).

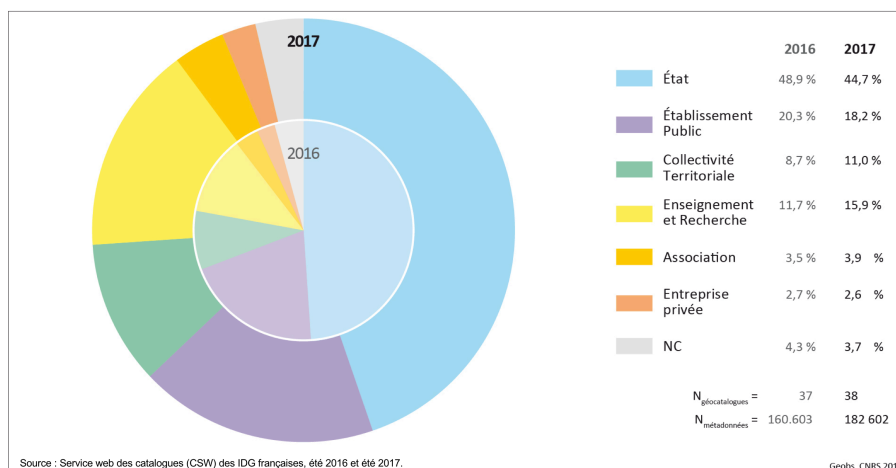


Figure 6. Évolution de la couverture organisationnelle des géocatalogues.

3.4. Mise en réseau des territoires : évolution de la couverture spatiale

La cartographie de l'emprise spatiale des données permet d'interroger l'égalité informationnelle des territoires (Gautreau et Noucher, 2016).

¹¹ La centralité d'intermédiation (*betweenness centrality*) C_B mesure la fréquence d'apparition d'un nœud sur les plus courts chemins entre les différents nœuds du réseau. $C_{B(2016)}=23,99$ $C_{B(2017)}=6,33$

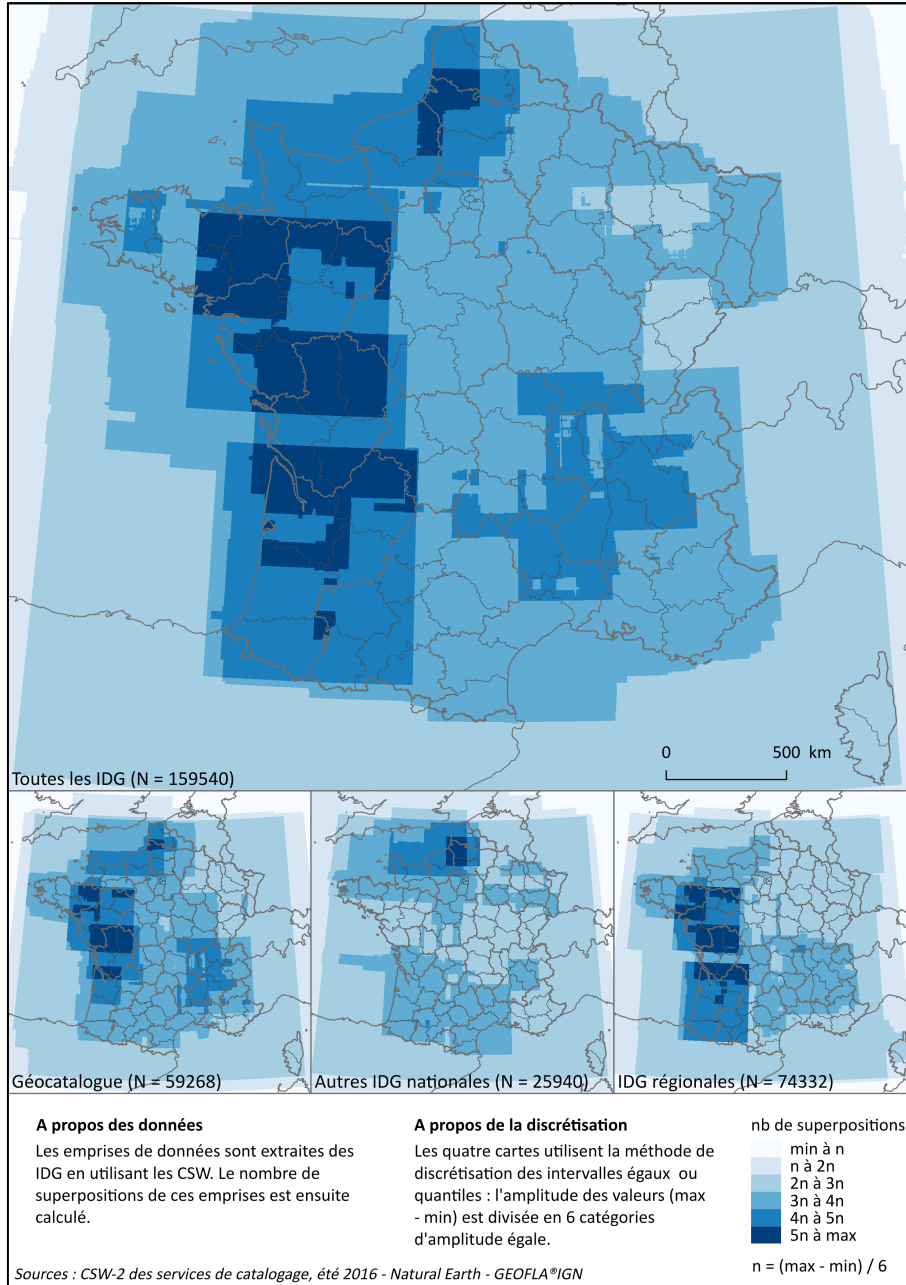


Figure 7. La couverture spatiale des métadonnées en 2016.

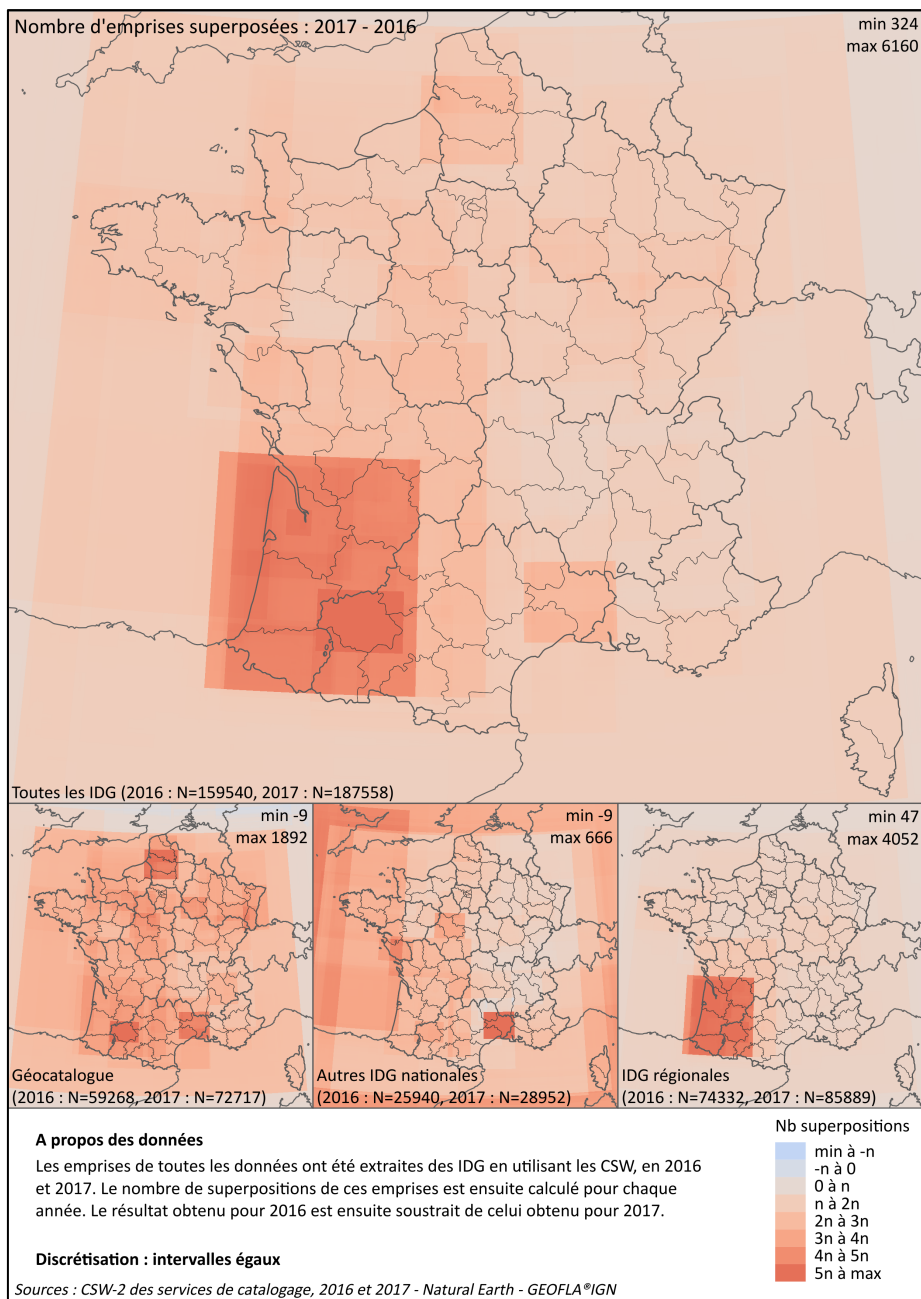


Figure 8. Evolution de la couverture spatiale des métadonnées entre 2016 et 2017.

Au-delà de l'identification ponctuelle des *hot-spots* informationnels qui peuvent être liés à des dynamiques locales (opération massive de numérisation ; territoires tests d'expérimentation ; etc.) ou au contraire des *zones blanches* en données (figure 7), il est aussi intéressant de suivre l'évolution des dynamiques de diffusion des données. La figure 8 présente ainsi le différentiel entre les couvertures spatiales des 160603 métadonnées recensées en 2016 et celles des 188602 métadonnées indexées¹² dans les mêmes géocatalogues en 2017. Si ce différentiel est quasi exclusivement positif, la discrétisation par quantile laisse entrevoir une certaine hétérogénéité dans l'augmentation du nombre de métadonnées. Les cartes présentées ici laissent ainsi apparaître l'importance de certaines initiatives départementales (Gers, Gard, Somme...) ou métropolitaines (Bordeaux) et la persistance des anciennes régions (l'Aquitaine plutôt que la Nouvelle Aquitaine) dans les dynamiques d'alimentation récente (sur 1 an) des infrastructures françaises de données géographiques.

4. Discussion

A partir de ces quatre indicateurs tests, nous avons expérimenté la pertinence d'une analyse diachronique du contenu des IDG. Un commentaire plus détaillé de chaque analyse thématique est disponible dans une précédente étude (Noucher et al., 2017). L'objet de la discussion est désormais d'envisager l'originalité mais aussi les atouts et limites de telles analyses au regard des travaux existants et dans la perspective des *Geographic Information Observatories* présentées en introduction.

Ainsi, le Ministère de la Transition écologique et solidaire utilise déjà le Géocatalogue national pour réaliser le rapportage¹³ exigé par la Commission européenne dans le cadre de la Directive INSPIRE (décision n°2009/442/CE). Il s'appuie sur le BRGM qui produit un suivi annuel et un rapport triennal. Le dernier rapport triennal, validé par le Conseil National de l'Information Géographique (CNIG), diffusé en juillet 2016, dresse le bilan de la période 2013-2015 et a été établi à partir d'une enquête, de retours d'utilisateurs et d'une série d'indicateurs. Ces derniers sont produits *via* l'analyse des séries de données et de services géographiques qui s'inscrivent dans le périmètre de la Directive et étaient disponibles au 31 décembre 2015 dans le géocatalogue national. Ils concernent essentiellement les suivis des métadonnées, des données, des services en réseaux et des actions de coordination. L'analyse de l'évolution des volumes (données, métadonnées, services) et de leur niveau de conformité permet de comparer entre elles les plateformes nationales et d'opérer leur suivi dans le temps. Le rapport de la France souligne ainsi le doublement annuel de métadonnées¹⁴ et précise que près de 90% d'entre-elles sont conformes à la réglementation européenne. Mais ce rapport ne fournit aucune analyse précise du patrimoine de données disponible, des champs thématiques prioritaires, des territoires les plus / les moins couverts, des échelles des données diffusées, de la répartition par

¹² Le nombre d'emprises traitées sur ces cartes est légèrement inférieur (159540 en 2016 et 187558 en 2017) au nombre de métadonnées indexées bien que plusieurs emprises puissent être associées à une métadonnée. Cette différence est liée au choix que nous avons fait, pour faciliter la diffusion des résultats, de ne centrer ces cartes que sur l'hexagone.

¹³ Rapport d'évaluation normé selon un modèle proposé par la Commission.

¹⁴ De 4026 métadonnées en 2013 à 34473 en 2016 soit une augmentation de 8,6 fois.

type d'organisation, des périodes de validité des données proposées. La décision de la Commission (2009/442/CE) du 5 juin 2009 inscrit le rapportage avant tout dans une logique de benchmarking¹⁵ visant à créer un référentiel cognitif commun destiné à rendre les données comparables d'un pays à l'autre, sans porter de réflexion de fond sur les types de contenus à créer ni sur leur pertinence. En ce sens, l'évaluation des IDG met ainsi les territoires « sous pression statistique » (Bruno et Didier, 2015) pour comparer, classer, hiérarchiser les types d'organisation ou « les territoires qui ne joueraient pas le jeu »¹⁶. Les tableaux de bord produit par la sphère institutionnelle semblent donc relever aujourd'hui davantage d'une « logique de contrôle et de stimulation des Etats en situant leurs performances sur des échelles normatives » (Alphandéry et al., 2012) que d'une volonté de mieux comprendre les connaissances territoriales qu'elles sont susceptibles de diffuser ou des flux informationnels qu'elles sont potentiellement en train de ré-agencer (Noucher, 2013).

Notre expérimentation, menée à partir de l'observation diachronique du contenu de 39 IDG françaises, nous a permis de tester, sur des volumes de données importants, l'intérêt de ce type d'analyse pour compléter le rapportage tel qu'il est réalisé actuellement en France. La poursuite de ce travail est envisagée pour engager des analyses diachroniques sur un pas de temps plus long.

Néanmoins, il est important de souligner trois limites à ce travail. Premièrement le caractère cumulatif des corpus disponibles en ligne interroge la manière de gérer les données obsolètes, les données « mortes » ou encore les données publiées par mégarde qui ont tendance à s'accumuler dans ces infrastructures rarement mises à jour. À ce stade, les observations diachroniques les agrègent sans pouvoir les distinguer, introduisant un biais dans les résultats. Deuxièmement, la faisabilité et la robustesse des analyses proposées sont liées aux caractères institutionnel et ouvert des IDG étudiées. Or, la mise en œuvre d'observatoires de l'information géographique circulant sur le Web au sens de Janowicz et al. (2014) nécessiterait d'analyser d'autres types de plateformes moins normées (de cartographie collaborative ou issues des réseaux sociaux) qui diffusent des données géographiques plus difficilement accessibles. Leur prise en compte pour dresser un tableau de bord qui tend vers l'exhaustif nécessiterait de démultiplier les chaînes d'extraction et d'archivage pour s'adapter à leur spécificité ou détourner leur restriction d'accès. Enfin troisièmement, la prise en compte des usages de ces plateformes est indispensable dans un contexte dynamique (Georis-Creuseveau et al., 2018). Elle constitue en effet une des clés de compréhension du fonctionnement et de l'évolution de ces IDG, assimilables à des systèmes complexes et adaptatifs (Grus et al., 2010).

5. Conclusion

Par cette contribution nous souhaitons démontrer l'intérêt de la mise en œuvre d'observatoires de l'information géographique circulant sur le Web pour appréhender

¹⁵ Le *benchmarking* est une technique issue du marketing pour comparer les performances des entreprises. Un *benchmark* est un indicateur chiffré de performance dans un domaine donné tiré de l'observation des résultats de l'entreprise qui a réussi le mieux dans ce domaine. Cet indicateur peut servir à définir les objectifs de l'entreprise qui cherche à rivaliser avec elle.

¹⁶ Expression du correspondant INSPIRE pour la France relevée lors d'un entretien (Paris, La Défense, le 27 mai 2016).

les dynamiques qui affectent le capital informationnel des territoires et les fluctuations de sa répartition (thématique, spatiale, organisationnelle) dans une logique de compréhension et non d'évaluation. Au-delà de la seule composante des données, l'observation des contenus informationnels de ces IDG, leur qualification et évolution, il apparaît clairement que les composantes organisationnelles et les distributions spatiales sont des dimensions additionnelles fondamentales pour la mise en œuvre d'outils d'observation élaborés. C'est effectivement un premier pas qui nous semble important pour la conception d'outils d'observations élaborés prenant en compte non seulement la composante informationnelle mais aussi les pratiques et les utilisations. De telles démarches imposeront le développement de techniques d'observation opérationnelles qui passeront par la mise en œuvre de protocoles de caractérisation des méthodes et outils mis en œuvre dans les traitements de données géographiques, ce qui reste un challenge important au vu des complexités informationnelles sous-jacentes.

L'approfondissement de l'analyse du patrimoine informationnel et des usages des IDG ainsi que la diversification des types d'infrastructures web à étudier représentent ainsi des défis de recherche pour opérationnaliser le concept d'observatoire de l'information géographique.

Annexe 1 : Liste et URL des IDG étudiées (opérationnelles au printemps 2019)

ID	NOM	ECHELON	URL
1	Agroenvgeo	National	https://agroenvgeo.data.inra.fr/
2	Atlasante	National	https://www.atlasante.fr/accueil
3	CARTOMER	National	http://cartographie.aires-marines.fr/
4	Cartorisque Géorisques	National	http://www.georisques.gouv.fr/
5	Data SHOM	National	https://data.shom.fr/
6	Geolittoral	National	http://www.geolittoral.developpement-durable.gouv.fr/
7	Géoportail de l'urbanisme	National	https://www.geoportail-urbanisme.gouv.fr/
8	Géocatalogue	National	http://www.geocatalogue.fr
9	GEOSUD	National	http://ids.equipep-geosud.fr/
10	InfoTerre	National	http://infoterre.brgm.fr/
11	ONML	National	http://www.onml.fr/accueil/
12	Sextant	National	https://sextant.ifremer.fr/fr/accueil
13	SIE EauFrance	National	https://www.eaufrance.fr/
14	Geo-IDE	National	http://catalogue.geo-ide.developpement-durable.gouv.fr/
15	GUYANE SIG	Régional	http://www.guyane-sig.fr/
16	KaruGéo	Régional	https://www.karugeo.fr/
17	CIGAL devenu GéoGrandEst	Régional	https://www.geograndest.fr/
18	CRAIG	Régional	https://www.craig.fr/
19	CRIGE PACA	Régional	http://www.crige-paca.org/
20	GéoBourgogne devenu IDéOBFC	Régional	https://www.ideobfc.fr/
21	GeoBretagne	Régional	https://cms.geobretagne.fr/
22	GeoGuyane	Régional	https://www.geoguyane.fr/accueil
23	GeoMartinique	Régional	http://www.geomartinique.fr/accueil
24	GeoMayotte	Régional	http://www.geomayotte.fr/accueil
25	GeoNormandie	Régional	https://www.geonormandie.fr/accueil
26	GEOPAL	Régional	https://www.geopal.org/accueil
27	GeoPicardie	Régional	https://www.geopicardie.fr/portail/
28	GEOREP	Régional	https://georep.nc/
29	GEORHONEALPES	Régional	Fusionné avec CRAIG
30	MiPyGeo	Régional	Fusionné avec SIG LR
31	PEGASE	Régional	Fusionné avec PIGMA
32	PEIGEO	Régional	http://peigeo.re/
33	PIGMA	Régional	https://portail.pigma.org/
34	PPIGE	Régional	https://www.ppige-npdc.fr/portail/
35	SIG L-R devenu OpenIG	Régional	http://amigo.siglr.org/
36	SIGERIF	Régional	Pas de site web
37	SIGLOIRE	Régional	http://www.sigloire.fr/
38	SIGOGNE	Régional	https://www.sigogne.org/
39	SIGRS GeoRhena	Régional	http://www.georhena.eu/

Annexe 2 : matériel de recherche complémentaire

Le site <http://geobs.cnrs.fr> permet d'accéder à la liste des services web de catalogage des infrastructures de données géographiques étudiées, au script conçu pour extraire les métadonnées, aux données exploitées pour la production des indicateurs présentées dans cet article ainsi qu'à quelques (géo)visualisations interactives.

Remerciements

Le projet GÉOBS, qui a supporté ce travail, a été soutenu financièrement par le CNRS et la Région Nouvelle-Aquitaine (Bourse 2014-1R40103). Les auteurs tiennent à remercier les membres de l'équipe, en particulier Adeline Maulpoix qui a mis en œuvre l'enquête auprès des coordinateurs des IDG françaises, Julie Pierson et Mathias Rouan qui ont développé la chaîne d'extraction des métadonnées, ainsi qu'Olivier Pissoat et Grégoire Le Campion qui ont contribué à la réalisation des figures.

Bibliographie

- Adams B., Gahegan M., Gupta P., Hosking R., (2014). Geographic Information Observatories for supporting Sciences. In Proceedings of Workshop on Geographic Information Observatories 2014, <http://ceur-ws.org/Vol-1273/>, 32-39.
- Alphandéry P., Fortier A., Sourdril A. (2012). Les données entre normalisation et territoire : la construction de la trame verte et bleue. Développement durable et territoires, 3/2.
- Ballatore A. (2014). Exploring the geographic information universe: the role of search technologies. In Proceedings of Workshop on Geographic Information Observatories 2014, <http://ceur-ws.org/Vol-1273/>, 23-28.
- Bruno I., Didier E. (2015). *Benchmarking, l'Etat sous pression statistique*. La Découverte.
- Gautreau P., Noucher M. (2016). Information géographique numérique et justice spatiale : les promesses du partage.. Justice Spatiale / Spatial Justice, 10, <http://www.jssj.org>
- Georis-Creuseveau J., Claramunt C., Gourmelon F., Pinaud B., David L. (2018). A diachronic perspective on the use of french Spatial Data Infrastructure. Journal of Geographic Information System <http://www.scirp.org/journal/jgis/>, 10, 344-361, <https://doi.org/10.4236/jgis.2018.104018>
- Grus L., Cromptvoets J., Bregt A.K. (2010). Spatial data infrastructures as complex adaptive systems. International Journal of Geographical Information Science, 24, 439-463.
- Janowicz K., Adams B., McKenzie G., Kauppinen T. (2014). Towards geographic information observatories. In Proceedings of Workshop on Geographic Information Observatories 2014, <http://ceur-ws.org/Vol-1273/>, 1-5.
- Libourel T., Passouant M., Loireau M. (2009). Introduction du chapitre « Systèmes d'observation et observatoires ». Dossier Agropolis International , 9.
- Loveluck B. (2015). *Réseaux, libertés et contrôle. Une généalogie politique d'Internet*. Armand Colin.

- Masser, I. (2010). *Building European Spatial Data Infrastructures*. ESRI Press: Redlands, CA, USA.
- McKenzie G., Janowicz K., Gao S., Yang J., Hu Y. (2015). POI Pulse: A Multi-Granular, Semantic Signatures-Based Information Observatory for the Interactive Visualization of Big Geosocial Data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 50/ 2, 71-85.
- Miller H. J. (2017). Geographic information observatories and opportunistic GIScience. *Progress in Human Geography*, 41(4), 489-500.
- Noucher M., Gourmelon F., Gautreau P., Georis-Creuseveau J., Maulpoix A., Pierson J., Pinède N., Pissot O. et Rouan M. (2017). Spatial Data Sharing: A Pilot Study of French SDIs. *ISPRS International Journal Geographic Information*, 6, 99.
- Noucher M. (2013). Infrastructures de données géographiques et flux d'information environnementale. *Netcom*, 27-1/2, 120-147.
- De Sède-Marceau M.-H., François S., Pauc B. (2018). OPTEER, une expérience de construction collective d'un dispositif de connaissance et d'analyse territoriale par et pour les acteurs de la transition énergétique. *Revue Internationale de Géomatique*, 28/1, 95-124.
- Rajabifard A., Feeney M.-E., Williamson I. et Masser, I. (2003). Chapter 6, National SDI Initiatives. In Williamson, I., Rajabifard, A., Feeney, M.-E.F., (Éds). *Development of Spatial Data Infrastructures: From Concept to Reality*. Taylor & Francis: London, UK, 95–109.
- Rouan M. et Pierson J. (2016). CSW_Harvester : une chaîne d'extraction et d'analyse d'informations à partir des services web de catalogage (CSW) des Infrastructures de Données Géographiques nationales et régionales en France. Atelier - EXtraction de Connaissances à partir de données Spatialisées (EXCES) – SAGEO'2016. Nice, France.