GÉOBS :
LES INFRASTRUCTURES DE DONNEES GEOGRAPHIQUES DANS LA GOUVERNANCE INFORMATIONNELLE DE L'ENVIRONNEMENT

PROJET DE RECHERCHE FINANCE PAR LA REGION NOUVELLE AQUITAINE (2015 – 2017)
COORDONNE PAR LE CNRS (LABORATOIRES PASSAGES-BORDEAUX ET LETG-BREST)

Analyse des géocatalogues des IDG françaises : note méthodologique

UMR Passages — BORDEAUX
UMR LETG — BREST
UMR PRODIG — PARIS
UMR LaBRI — BORDEAUX
EA MICA — BORDEAUX

Analyse des géocatalogues des IDG françaises : note méthodologique	
Date de création	17 octobre 2016
Dernière révision	24 octobre 2016
Auteurs	- Mathias Rouan, CNRS, UMR LETG, Brest - Julie Pierson, CNRS, UMR Passages, Bordeaux Campus - Matthieu Noucher, CNRS, UMR Passages, Bordeaux Campus - Françoise Gourmelon, CNRS, UMR LETG, Brest
Licence	Ce document est mis à disposition selon les termes de la Licence Ouverte
Citer le document	Rouan M., Pierson J., Noucher M., Gourmelon F., « Analyse des Géocatalogues des Infrastructures de Données Géographiques en France : note méthodologique ». Rapport intermédiaire du projet de recherche GÉOBS. 2016. 18 p.

SOMMAIRE

LE PROJET DE RECHERCHE GÉOBS

GÉOBS est un projet de recherche coordonné par le CNRS (laboratoires Passages-Bordeaux et LETG-Brest) et financé par la région Nouvelle Aquitaine sur la période 2015-2017. Son objectif est d'étudier les flux d'information géographique qui circulent sur le web pour analyser les stratégies des pouvoirs publics afin d'organiser la **circulation des connaissances sur l'environnement**. L'analyse des **contenus** et des **usages** de l'information géographique institutionnelle opérée par une **observation multi-niveaux** (du national au local) des **IDG** est l'enjeu scientifique du projet de manière à comprendre les stratégies contemporaines de « **gouvernance informationnelle** » de l'environnement. **GÉOBS** entend également répondre à un enjeu institutionnel dans la mesure où certains acteurs internationaux, régionaux et locaux chargés de produire de l'information géographique se demandent encore aujourd'hui quel est l'impact de ces dispositifs socio-techniques en termes d'amélioration de la gestion des territoires.

Trois axes de recherche sont menés :

- analyse du **contenu** des IDG - par la mobilisation d'une démarche interdisciplinaire associant géographie, informatique et sciences de l'information et de la communication, GÉOBS étudie les sites web et les géocatalogues pour décrypter les stratégies d'affichage et la couverture organisationnelle, thématique, spatiale, temporelle des données diffusées,
- analyse des **stratégies** des promoteurs des IDG - à partir d'enquêtes et d'entretiens menés avec plusieurs promoteurs d'IDG et d'autres plateformes qui diffusent de l'information géographique (observatoires, portails *opendata*, etc.), GÉOBS retrace l'origine, l'état actuel et les perspectives d'évolution des IDG face aux nouvelles offres en matière de production/diffusion d'information géographique,
- analyse des **usages** des IDG - une enquête nationale complétée par des études de cas est réalisée pour comprendre l'impact de ces dispositifs sur les pratiques quotidiennes de gestion des territoires.

Pour mener à bien cette triple analyse, le projet réunit des chercheurs rattachés à 3 laboratoires de géographie (Passages-Bordeaux, LETG-Brest, PRODIG-Paris), un laboratoire d'informatique (LaBRI-Bordeaux) et un laboratoire en sciences de l'information et de la communication (MICA-Bordeaux).

La finalité du projet est de mettre en place un **prototype d'observatoire** des IDG qui permette de dresser un panorama dynamique des flux d'information géographique sur l'environnement et d'analyser leur contribution aux politiques de gestion des territoires.

Site web du projet : http://www-iuem.univ-brest.fr/pops/projects/geobs
Contacts : matthieu.noucher@cnrs.fr // francoise.gourmelon@univ-brest.fr

OBJECTIF DE LA NOTE METHODOLOGIQUE

Cette note de cadrage précise la **méthode d'analyse des géocatalogues** mise en pratique par l'équipe GÉOBS pour explorer les métadonnées des infrastructures de données géographiques nationales et régionales qui sont étudiées par le projet.

Elle ne concerne donc qu'une partie du projet (le volet « catalogage » de l'axe « analyse de contenu »). Par ailleurs, il ne s'agit pas d'un article scientifique ou d'un livrable présentant des résultats statistiques mais d'une annexe méthodologique qui renseigne le protocole d'extraction, d'archivage, d'analyse et de visualisation des métadonnées.

Cette note méthodologique publiée sur l'archive ouverte institutionnelle HAL-SHS vient donc compléter un ensemble d'autres documents relatifs au projet et accessibles sur Internet. En effet, dans le cadre d'une démarche en cohérence avec son objet d'observation et avec les principes de l'*Open Science*, les productions de GÉOBS sont mises à disposition tout au long du projet sous la Licence Ouverte d'Etalab. Plusieurs canaux de diffusion sont mobilisés pour diffuser les résultats du projet :

- Les rapports intermédiaires, comme cette note de cadrage, sont diffusés sur le site web collaboratif du projet[1],
- Les visualisations interactives sont diffusées sur une plateforme web rassemblant plusieurs travaux de recherche qui décryptent le géoweb[2],
- A chaque visualisation en ligne sont associés les jeux de données diffusés sur data.gouv.fr[3],
- Les scripts développés dans le cadre du programme sont diffusés sur la plateforme d'hébergement et de gestion de développement de logiciels *github*[4],
- Enfin, les versions auteur (dites « pre-print ») des publications scientifiques dans des revues en *open access,* sont déposées sur les dépôts institutionnels HAL-SHS.

[1] http://www-iuem.univ-brest.fr/pops/projects/geobs
[2] http://www.geobs.cnrs.fr
[3] https://www.data.gouv.fr/fr/organizations/umr-5319-passages/#datasets
[4] https://github.com/LETG/csw-harvester

La notion d'Infrastructure de Données Géographiques

Nous adoptons la définition de Rajabifard *et al.* (2002)[5] reprise par une large communauté internationale qui s'intéresse à l'univers de l'information géographique, pour définir les Infrastructures de Données Géographiques (IDG) comme **des solutions fédérées qui rassemblent, dans un cadre dynamique, les informations, les réseaux informatiques, les normes et standards, les accords organisationnels et les ressources humaines nécessaires pour faciliter et coordonner le partage, l'accès et la gestion des informations géographiques**.

Complémentaire de cette définition fondée sur les composantes, des approches insistant sur la logique de réseaux des IDG conduisent à les représenter comme des « **organisations qui produisent, utilisent et partagent des informations géographiques, et en termes de flux entre ces organisations. Organisation et flux forment un réseau de partage et d'échange d'informations** » (Vandenbroucke et al., 2009) [6].

Le corpus de GÉOBS est composé de **45 IDG** identifiées à partir d'un premier inventaire réalisé en 2014 par l'AFIGEO. L'actualisation de cet inventaire a été réalisée à partir de l'exploration des sites web (vérification de leur maintien) et de plusieurs entretiens auprès des animateurs du réseau national des IDG. Depuis l'inventaire de 2014, certaines IDG ont disparu, fusionné ou ont été créées ce qui témoigne d'un contexte institutionnel encore instable.

Fort de ce travail de recension, **16 dispositifs nationaux et 29 régionaux** composent le corpus de GÉOBS. Ils ont en commun :

- d'être en adéquation avec la définition académique des IDG ;
- de se revendiquer de cette notion soit en s'affichant dans l'inventaire de l'AFIGEO, soit en faisant référence aux IDG ou aux textes de cadrage qui y sont associées.

Le projet de recherche GÉOBS ne cherche pas à établir une définition stricte des IDG mais au contraire à comprendre les différentes formes d'appropriation de cette notion en étudiant des dispositifs qui l'ont, à un moment ou à un autre, revendiquée.

Analyser les métadonnées pour décrypter le contenu des IDG

Les métadonnées correspondent à un ensemble structuré d'informations décrivant une ressource telle qu'un jeu de données géographiques, une carte ou un document… Les métadonnées relatives aux données géographiques concernent l'identification, l'étendue, la qualité, les aspects spatiaux et temporels, le contenu, la référence spatiale, la représentation des données, la distribution et d'autres propriétés utiles pour favoriser leur recherche et leur réutilisation.

La norme ISO 19115 -1:2014[7] définit le schéma requis pour décrire des informations géographiques et des services au moyen des métadonnées. Celles-ci sont généralement

[5] Rajabifard, A., Feeney, M.-E., and Williamson, I.P., 2002. Future directions for SDI development. *International Journal of Applied Earth Observation and Geoinformation*, 4 (1), 11–22. doi:10.1016/ S0303-2434(02)00002-8

[6] Vandenbroucke D., J. Crompvoets, G. Vancauwenberghe, E. Dessers, J. Van Orshoven, 2009, A Network Perspective on Spatial Data Infrastructures: Application to the Sub-national SDI of Flanders (Belgium). *Transactions in GIS*, 13,pp. 105-122

[7] http://www.iso.org/iso/fr/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=53798

regroupées dans un catalogue (ou **géocatalogue**) accessible sur le web et proposant un certain nombre de services sur les données géographiques : recherche, consultation, téléchargement et transformation. Pour mettre en œuvre ces services, les catalogues s'appuient sur des standards de l'OGC[8] comme le *Web Map Service* (WMS) pour la consultation, le *Web Feature Service* (WFS) et *Web Coverage Service* (WCS) pour le téléchargement ou le *Web Processing Service* (WPS) pour la transformation.

Notre analyse du contenu des géocatalogues repose sur le ***Catalog Service for the Web* (CSW)**. Ce standard permet de mettre en œuvre les services de découverte, recherche et moissonnage des métadonnées. Comme tous les services web de l'OGC, le CSW propose un certain nombre d'opérations interrogeables à travers le protocole HTTP qui permettent de construire des requêtes dont les résultats sont retournés en XML.

Parmi ces opérations on trouve notamment :

- *GetCapabilities* qui permet de récupérer les métadonnées du service (description, opérations, critères de recherche, standards) ;
- *GetRecords* qui permet de réaliser une recherche de métadonnées dans le catalogue et de récupérer les informations de métadonnées répondant aux critères.

Les IDG étudiées par GÉOBS et leur CSW

Les services de catalogues des 45 IDG du corpus ont été testés entre le 22 juin et le 22 août 2016. Les adresses (URL) de ces services de catalogage ont été inventoriées à partir de l'outil de monitoring[9] développé par la Mission de l'Information Géographique (MIG) du MEDDE ou directement demandées auprès des coordinateurs des IDG.

Sur ces 45 IDG, **37 services de catalogage opérationnels** ont pu être utilisés (tableau 1). C'est l'ensemble des métadonnées proposées à travers ces 37 CSW qui est analysé, le périmètre de GEOBS ne se limitant pas à un périmètre réglementaire ou thématique : l'objectif est d'analyser ce que les coordinateurs et membres des IDG rendent effectivement visible à travers leur catalogue de données géographiques.

[8] L'Open Geospatial Consortium est un groupement international qui vise à développer et promouvoir des standards ouverts afin de garantir l'interopérabilité des contenus, des services et des échanges dans les domaines de la géomatique : http://www.opengeospatial.org/

[9] http://geocat.docinspire.eu/geocats.php

Tableau 1. Liste des IDG retenues pour l'analyse et des CSW-2 opérationnels associés.

ID	NOM	ECHELON	URL CSW-2 (valide à l'été 2016)
01	Agroenvgeo	National	https://agroenvgeo.data.inra.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
02	Atlasante	National	http://www.atlasante.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
03	CARTOMER	National	http://cartographie.aires-marines.fr/geosource/srv/fre/csw?service=CSW&request=GetCapabilities
04	Carto/Géorisques	National	http://www.mongeosource.fr/geosource/1033/fre/csw?version=2.0.2&REQUEST=GetCapabilities
05	Data SHOM	National	http://services.data.shom.fr/csw/ISOAP?service=CSW&request=GetCapabilities
06	GeoFoncier	National	/
07	Geolittoral	National	http://www.mongeosource.fr/geosource/1111/fre/csw?service=CSW&request=GetCapabilities
08	GeoCatalogue	National	http://www.geocatalogue.fr/api-public/servicesRest?SERVICE=CSW&request=getCapabilities
09	GEOSUD	National	http://geosud.ign.fr/csw/ISOAP?service=CSW&request=GetCapabilities
10	ONML	National	/
11	Sextant	National	http://sextant.ifremer.fr/geonetwork/srv/eng/csw?service=CSW&request=GetCapabilities
12	SIE EauFrance	National	http://www.data.eaufrance.fr/geosource/srv/fre/csw?service=CSW&request=GetCapabilities
13	SINP	National	http://inventaire.naturefrance.fr/geosource/srv/fre/csw?service=CSW&request=GetCapabilities
14	InfoTerre	National	Via le géocatalogue national
15	Géoportail Urbanisme	National	Via le géocatalogue national
16	Géo-IDE	National	http://catalogue.geo-ide.developpement-durable.gouv.fr/catalogue/srv/eng/csw-moissonnable-ds-2?service=CSW&request=GetCapabilities
17	APUR	Régional	Via le géocatalogue national
18	CIGAL	Régional	https://www.cigalsace.org/geonetwork/srv/eng/csw?service=CSW&request=GetCapabilities
19	CIGEO	Régional	http://infogeo.ct-corse.fr/geoportal/csw/discovery?service=CSW&request=GetCapabilities
20	CRAIG	Régional	http://ids.craig.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
21	CRIGE PACA	Régional	http://geocatalogue.crige-paca.org/geonetwork/srv/fr/csw?service=CSW&request=GetCapabilities
22	GeoBourgogne	Régional	http://catalogue.geobourgogne.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
23	GeoBretagne	Régional	http://geobretagne.fr/geonetwork/srv/eng/csw?service=CSW&request=GetCapabilities
24	Geo-Centre	Régional	http://catalogue.geo-centre.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
25	GeoGuyane	Régional	http://catalogue.geoguyane.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
26	GeoLimousin	Régional	http://catalogue.geolimousin.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
27	GeoMartinique	Régional	http://www.geomartinique.fr/geonetwork/srv/fre/csw
28	GeoMayotte	Régional	http://www.geomayotte.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
29	GeoNormandie	Régional	http://catalogue.geonormandie.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
30	GEOPAL	Régional	http://www.geopal.org/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
31	GeoPicardie	Régional	http://www.geopicardie.fr/geonetwork/srv/eng/csw-for-harvesters?service=CSW&request=GetCapabilities
32	GEOREP	Régional	http://www.geoportal.gouv.nc/geoportal/csw?service=CSW&request=GetCapabilities
33	GEORHONEALPES	Régional	http://catalogue.georhonealpes.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
34	PEGASE	Régional	http://catalogue.pegase-poitou-charentes.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
35	PEIGEO	Régional	http://www.peigeo.re/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
36	PIGMA	Régional	http://ids.pigma.org/geonetwork/srv/eng/csw?service=CSW&request=GetCapabilities
37	PPIGE	Régional	http://api.isogeo.com/services/ows/s/3f5cb018fe4e48dcac7f6039acd01962/c/cd8aeeede5324b54900264d6c5987d04/ZKyGZvBoeonP1vIzCeAAwJqDc4HifEMRC2fH4sTg4dRh6cy4A5G9JC?service=CSW&request=GetCapabilities
38	SIG L-R	Régional	http://geocatalogue.siglr.org/geonetwork/srv/fr/csw?service=CSW&request=GetCapabilities
39	SIG Pyrénées	Régional	/
40	SIGERIF	Régional	/
41	SIGLOIRE	Régional	http://catalogue.sigloire.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
42	SIGOGNE	Régional	http://www.sigogne.org:8080/geosource/srv/fre/csw?service=CSW&request=GetCapabilities
43	SIGRS	Régional	/
44	MiPyGéo	Régional	http://catalogue.mipygeo.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
45	Guyane-SIG	Régional	https://catalogue.guyane-sig.fr/geonetwork/srv/fr/csw?service=CSW&request=GetCapabilities

La liste des IDG recensées par le projet GEOBS ainsi que les URL vers leur site éditorial, portail de visualisation et catalogue de métadonnées sont disponibles à l'adresse suivante : http://geobs.cnrs.fr/#pages/portfolio/idg_inventaire.html

Extraction

Pour extraire les informations contenues dans les métadonnées des IDG, un script Python[10] 2.7 utilisant la bibliothèque OWSLib[11] a été développé (figure 1). Cette bibliothèque offre la possibilité d'interroger des catalogues en utilisant les différents standards OGC, dont le CSW dans sa version 2.0.2. D'autres bibliothèques proposent des fonctionnalités équivalentes, comme par exemple GDAL/OGR en Shell ou Geotools en Java. OWSLib a été choisie pour sa documentation complète et sa simplicité d'utilisation.

Pour cette phase d'extraction, plusieurs problèmes d'ordre technique ont été rencontrés : selon la plate-forme utilisée par l'IDG (GeoNetwork, Isogeo, Prodige, Amigo...), l'implémentation du CSW peut varier, une IDG peut ne plus être interrogeable (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 deux mois (du 22 juin au 22 août 2016) pour la 1ère expérimentation. D'autres problèmes inhérents aux modes de remplissage des fiches de métadonnées ont été identifiés : 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 par exemple des dates (de création, de publication, de 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 permet de récupérer les fichiers XML des métadonnées et de réaliser leur analyse avec des requêtes XQuery. Cependant le souhait de pouvoir 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 Psycopg[12], adaptateur PostgreSQL le plus utilisé en Python.

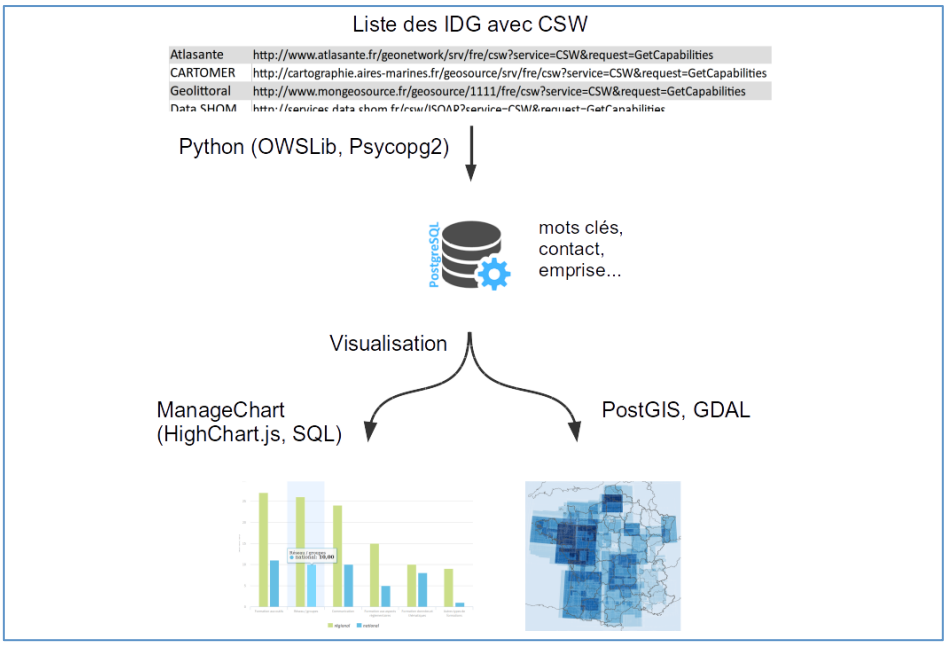

Figure 1. Schéma global de la chaîne de traitement avec les différentes technologies.

[10] https://github.com/LETG/csw-harvester
[11] https://github.com/geopython/OWSLib
[12] http://initd.org/psycopg/

Archivage

La norme ISO 19115 contient un très grand nombre de champs comme en témoigne ce diagramme UML (figure 2) décrivant uniquement la partie identification de la métadonnée.

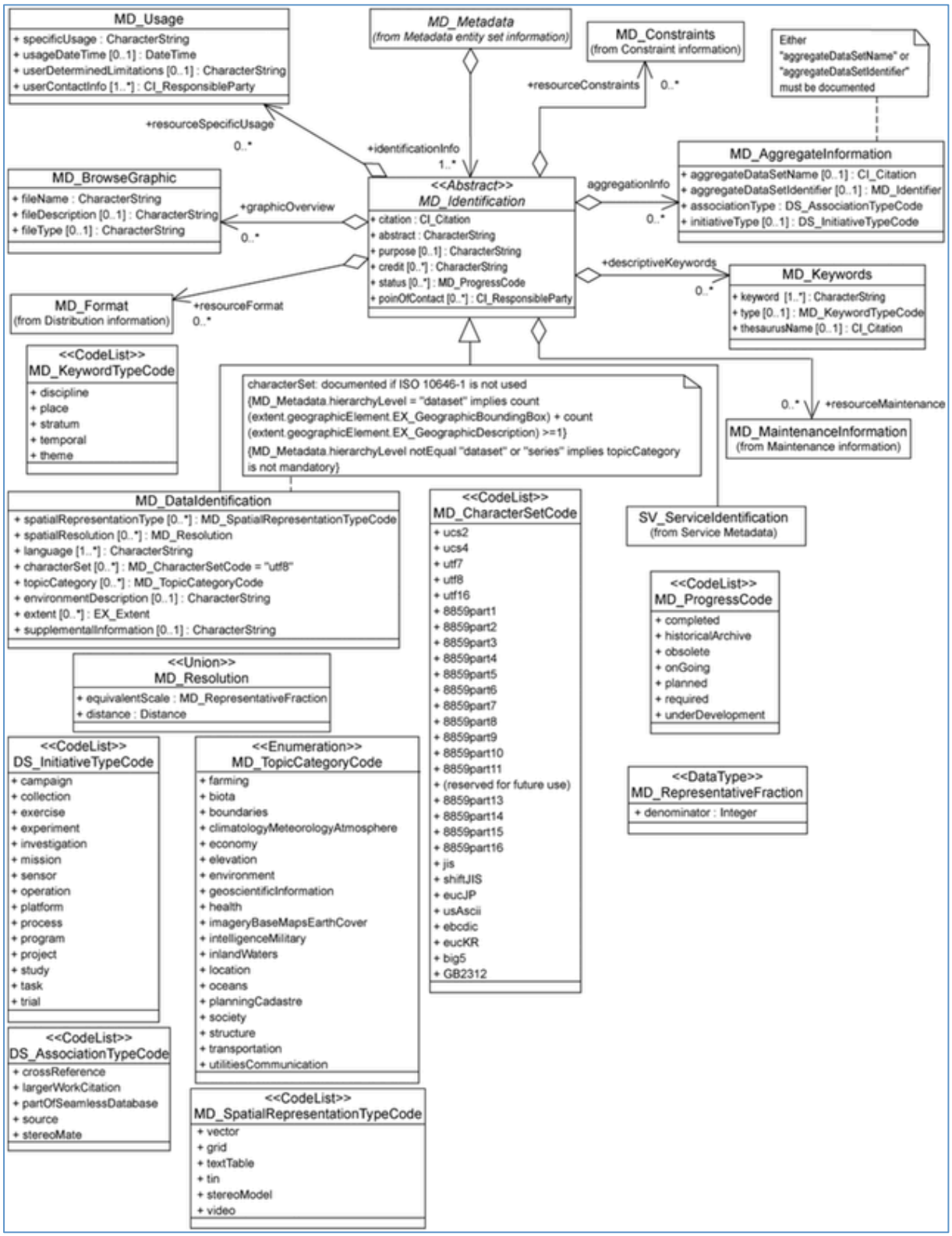

Figure 2. Diagramme UML « Identification Information » de la norme ISO 19115.

L'objectif de GÉOBS n'est pas d'explorer l'ensemble des métadonnées mais de les analyser au regard des quatre dimensions qui fondent les questions de recherche et alimente les autres chantiers du projet (concernant les stratégies des coordinateurs et les pratiques des usagers) à savoir :

1. le partage des données (accessibilité et utilisabilité),
2. la mise en réseau des outils (interopérabilité),
3. la mise en réseau des acteurs (géocollaboration),
4. la couverture territoriale des données (égalité informationnelle des territoires).

Il n'était donc pas question de proposer un schéma de base de données reprenant l'ensemble de la norme. Dans cette perspective, une sélection d'éléments pertinents a été réalisée et un modèle conceptuel élaboré (figure 3).

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 contient 3 entités (« Metadata », « DataIdentification » et « GeographicBoundingBox ») qui sont associées par des relations de cardinalité 1,1 ce qui occasionne une multiplication injustifiée du nombre de tables et donc une complexification des requêtes par l'ajout de jointure. Cependant cette structuration a l'avantage d'être évolutive en permettant l'ajout de nouveaux attributs ou entités/relations tout en conservant l'intégrité de la base sans modifier les requêtes existantes. 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.

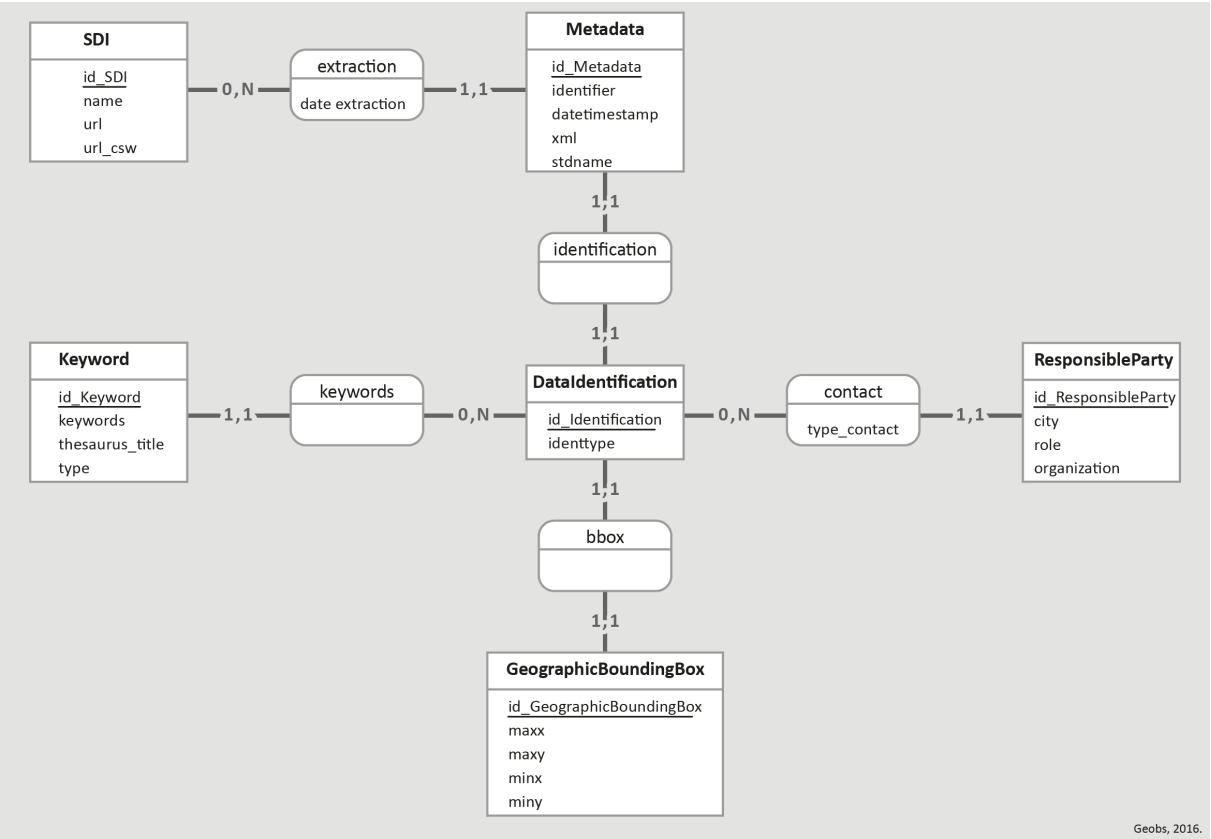

Figure 3. Modèle conceptuel de données.

La base de données a été implémentée sous PostgreSQL 9.3.

Analyse et visualisation

Dans le cadre de la mise en œuvre de l'IDG Indigeo[13] a été développée ManageChart, une application web de visualisation de graphiques dynamiques. Basée sur le framework PHP Symfony[14] et sur la librairie JavaScript Highcharts[15], ManageChart permet, à travers une interface graphique, de se connecter à des bases de données, d'en extraire des vues à l'aide de requêtes SQL et d'élaborer différents types de graphiques (courbes, histogrammes, diagrammes circulaires, polaires...). Ces graphiques sont ensuite générés dynamiquement à partir des requêtes exécutées à la volée (figure 4).

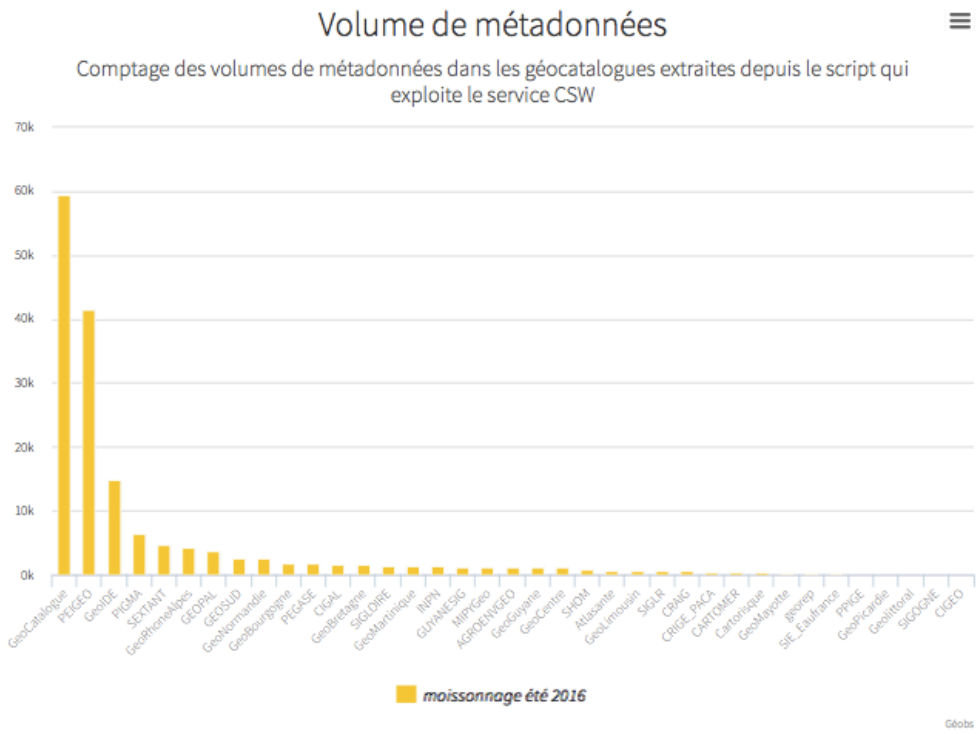

Figure 4. Graphique ManageChart représentant le volume de métadonnées.

Le choix des types de représentations (ligne, histogramme), couleurs et libellés du graphique se fait depuis une interface web de configuration, accessible avec un compte (figure 5).

[13] http://indigeo.fr
[14] https://symfony.com/
[15] http://www.highcharts.com/

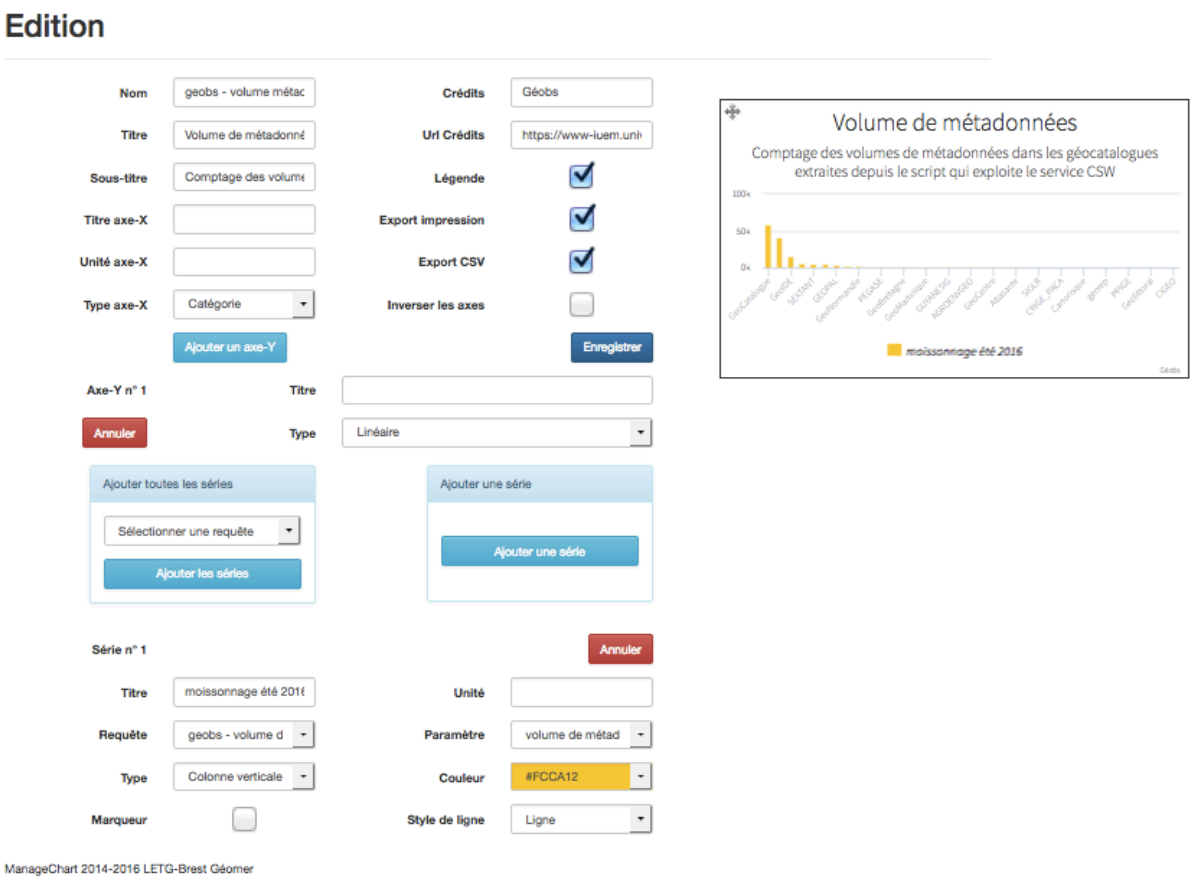

Figure 5. Interface de configuration dynamique d'un graphique avec ManageChart

Ces graphiques sont accessibles et paramétrables (taille et filtre de requête) depuis une URL[16]. Ce sont ces URL qui sont intégrées aux publications ou au site web de valorisation du projet : http://www.geobs.cnrs.fr

Les deux sections suivantes présentent deux exemples d'analyse réalisée sur les fiches de métadonnées à partir de la chaine qui vient d'être présentée.

[16] Exemple : https://www-iuem.univ-brest.fr/wapps/managechart/fr/chart/show/105/800/530/%7B%7BAttrSpatiaux%7D%7D

EXEMPLE N°1 : ACCESSIBILITE DES DONNEES CATALOGUEES

L'extraction des métadonnées issues des 37 services web de catalogage permet de dénombrer les données géographiques documentées, à l'été 2016. En plus du géocatalogue national qui compte 59 399 fiches de métadonnées soit un peu plus du tiers du corpus (160 603), on dénombre ainsi en moyenne 4300 fiches de métadonnées dans les 37 autres catalogues: 7200 en moyenne pour les catalogues nationaux et 3200 pour les régionaux.

Au-delà du comptage du nombre de métadonnées, nous avons cherché à calculer le nombre de données accessibles et exploitables (et non simplement indexées dans les géocatalogues). Pour y parvenir, trois niveaux d'analyse ont été mis en œuvre :

1. Une requête pour extraire les données qualifiées de « données ouvertes » ;
2. Une requête pour évaluer l'opérationnalité des liens (URL) associés aux données ouvertes ;
3. Une requête pour comparer les protocoles de distribution des données ouvertes.

Les données taggées « données ouvertes »

Un premier niveau d'analyse visait à identifier les métadonnées qualifiées avec le mot clé « données ouvertes » ou « open data ». Lors de la préparation de cette requête l'analyse du champ « keyword » de la table « keywords » a révélé une diversité de syntaxe des mots-clés « données ouvertes » et « open data » : avec ou sans accents, singulier/pluriel ou mélangé, ponctuation, coquille… mais également avec des concaténations de plusieurs mots clefs (tableau 2).

Tableau 2. Exemple de syntaxes différentes pour le mot clé « données ouvertes ».

commune ; bassin de vie ; donnée ouverte
donnees ouvertes
données ouvertes
donnée ouverte
donnée ouvertes
données ouvertes
données ouvertes ? en fonction du choix du producteur
données ouvertes,France Métropolitaine
données ouvertes.
données ouvertes;
donnéées ouvertes

Afin de prendre en compte cette hétérogénéité, le filtre de la requête a été affiné par l'utilisation des correspondances de motif ou pattern matching dans PostgreSQL :

```
(keywords LIKE '%donn%e%ouverte%' OR keywords LIKE '%open%data%')
```

Ici le signe pourcent (%) remplace toutes les chaînes de zéro ou plusieurs caractères.

Deux autres requêtes qui calculent respectivement les accessibilités moyennes pour 1) les IDG nationales ou régionales et 2) toutes les IDG, ont permis de finaliser le graphique présenté ci-après (figure 6).

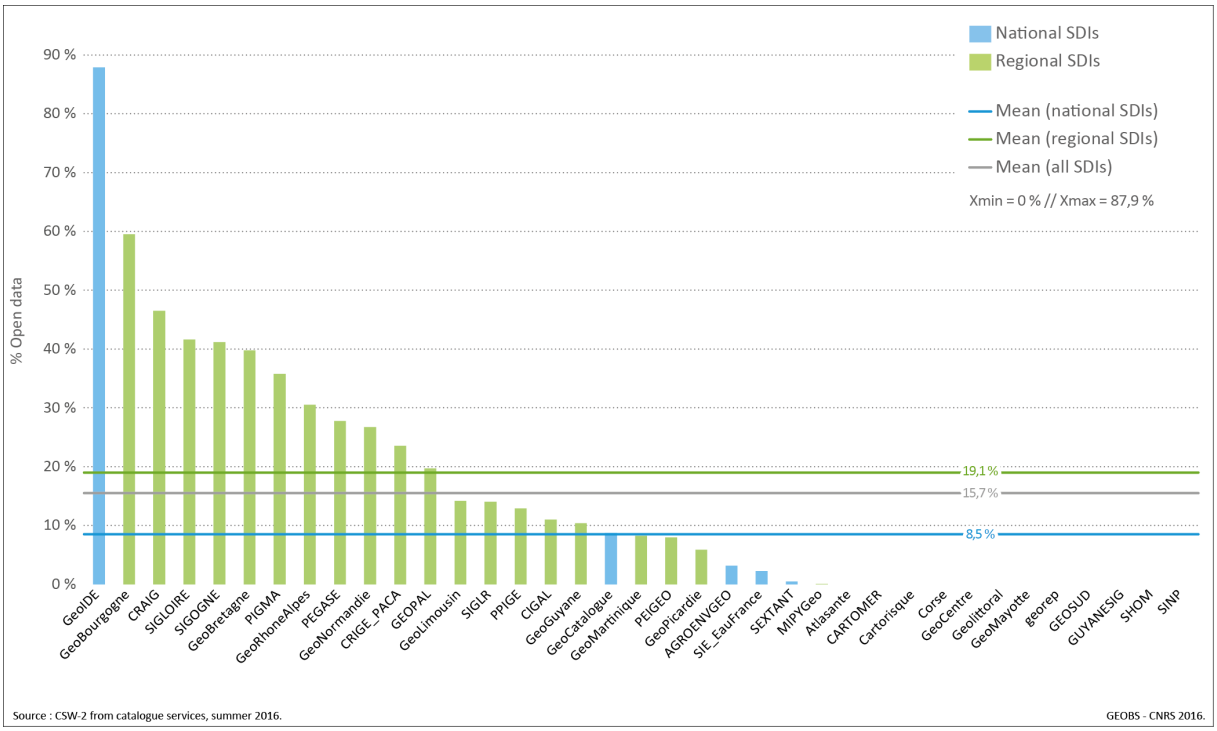

Figure 6. Part des données qualifiées de « données ouvertes » ou « open data » dans les géocatalogues.

L'opérationnalité des liens associés aux données « ouvertes »

L'analyse précédente est basée sur du déclaratif puisque seules les données qualifiées de données ouvertes ont été conservées pour établir les ratios. Pour aller plus loin dans l'analyse et, en particulier, pour évaluer l'opérationnalité de l'accès aux données, nous testons, à partir d'un script Python dédié, les liens associés aux données.

Trois types de réponses sont mis en évidence :

1. Le lien est valide ;
2. Le lien est invalide ;
3. Le lien ne s'ouvre pas au bout de 10 sec. (time-out).

Les résultats sont présentés ci-dessous et réduisent encore un peu la part des données accessibles (i.e. qualifiée de « données ouvertes » et dont le lien de téléchargement est opérationnel).

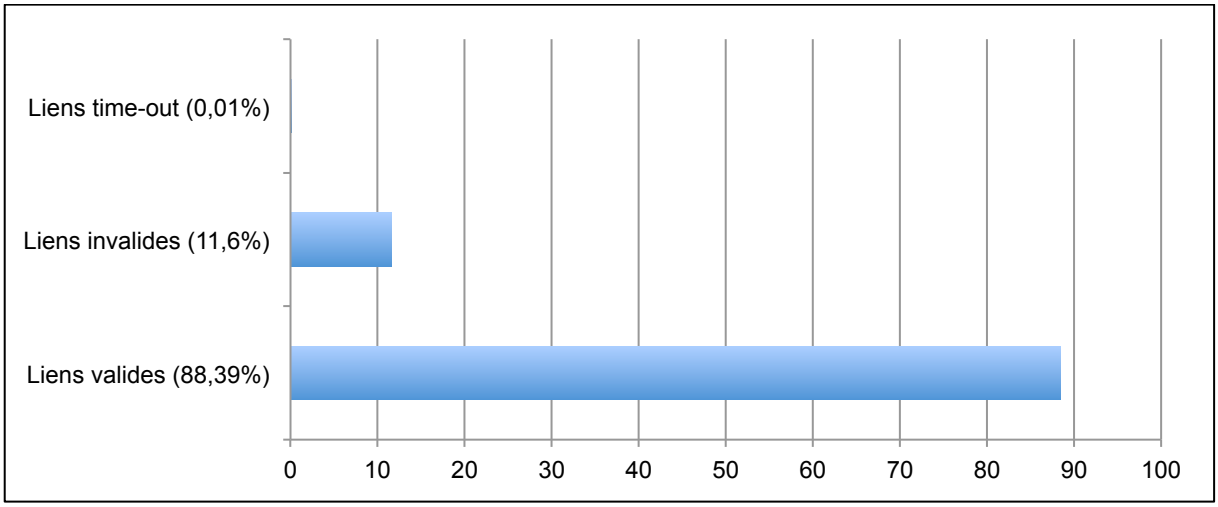

Figure 7. Opérationnalité des liens associés aux « données ouvertes » dans les géocatalogues.

Protocole de distribution

Enfin, une dernière requête permet d'identifier les types de protocole de distribution des liens opérationnels (étape 2) associés aux données ouvertes (étape 1). Cette requête utilise la balise protocol[17] et permet de distinguer des distributions d'images statiques (PNG, PDF), de services web (WMS, WFS), et de données dans des formats ouverts d'encodage comme le GeoJSON (figure 8).

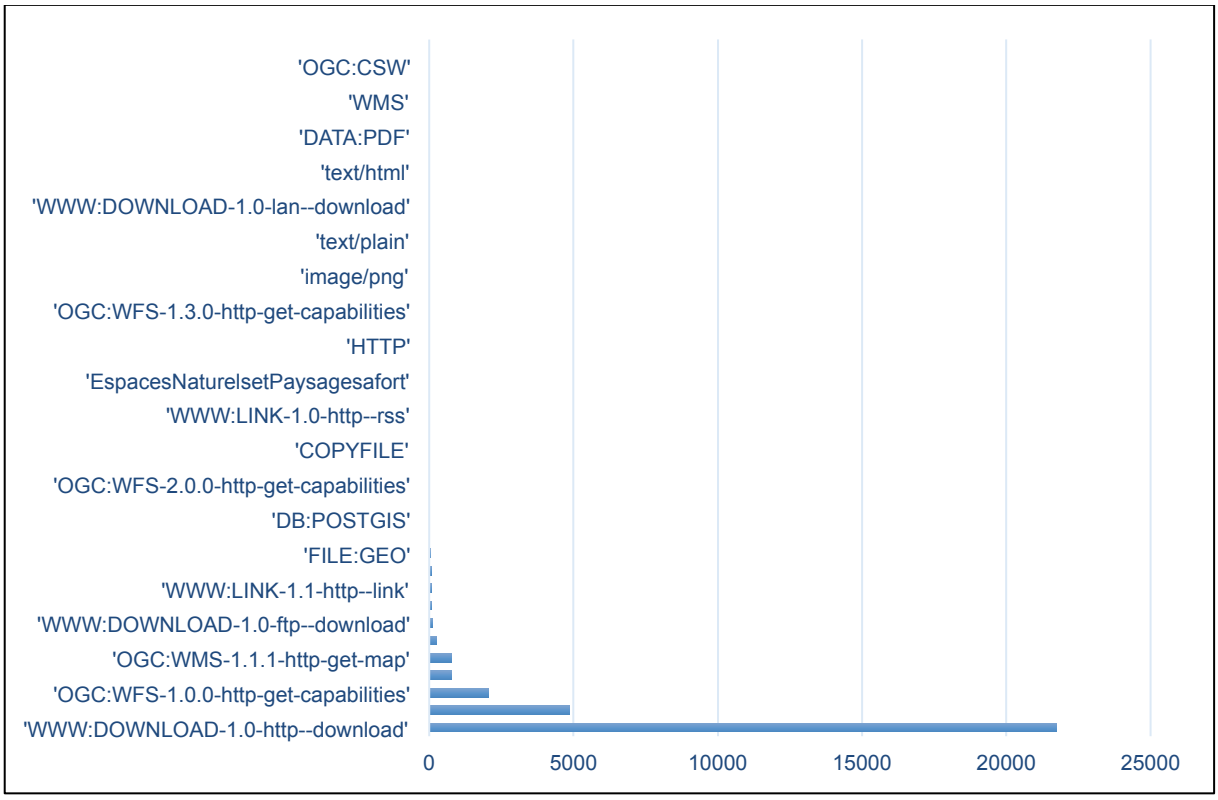

Figure 8. Type de distribution des liens opérationnels associés aux « données ouvertes » dans les géocatalogues.

[17] gmd:distributionInfo/gmd:MD_Distribution/gmd:transferOptions/gmd:MD_DigitalTransferOptions/gmd:onLine/gmd:CI_OnlineResource/gmd:protocol

EXEMPLE N°2 : COUVERTURE TERRITORIALE DES DONNEES

Cet exemple permet d'expliciter la méthode d'analyse de la couverture territoriale des données mise en œuvre par GÉOBS. Chaque métadonnée possède une ou plusieurs emprise(s) rectangulaire(s) définie(s) par quatre coordonnées est, ouest, nord et sud au sein de la balise <EX_GeographicBoundingBox>. Afin de visualiser ces emprises, une couche spatiale est créée à partir de la liste de ces coordonnées (figure 9) en utilisant l'extension PostGIS de PostgreSQL. Cette extension permet le support d'objets géographiques dans la base, et donc de l'utiliser comme une base d'informations géographiques (SIG).

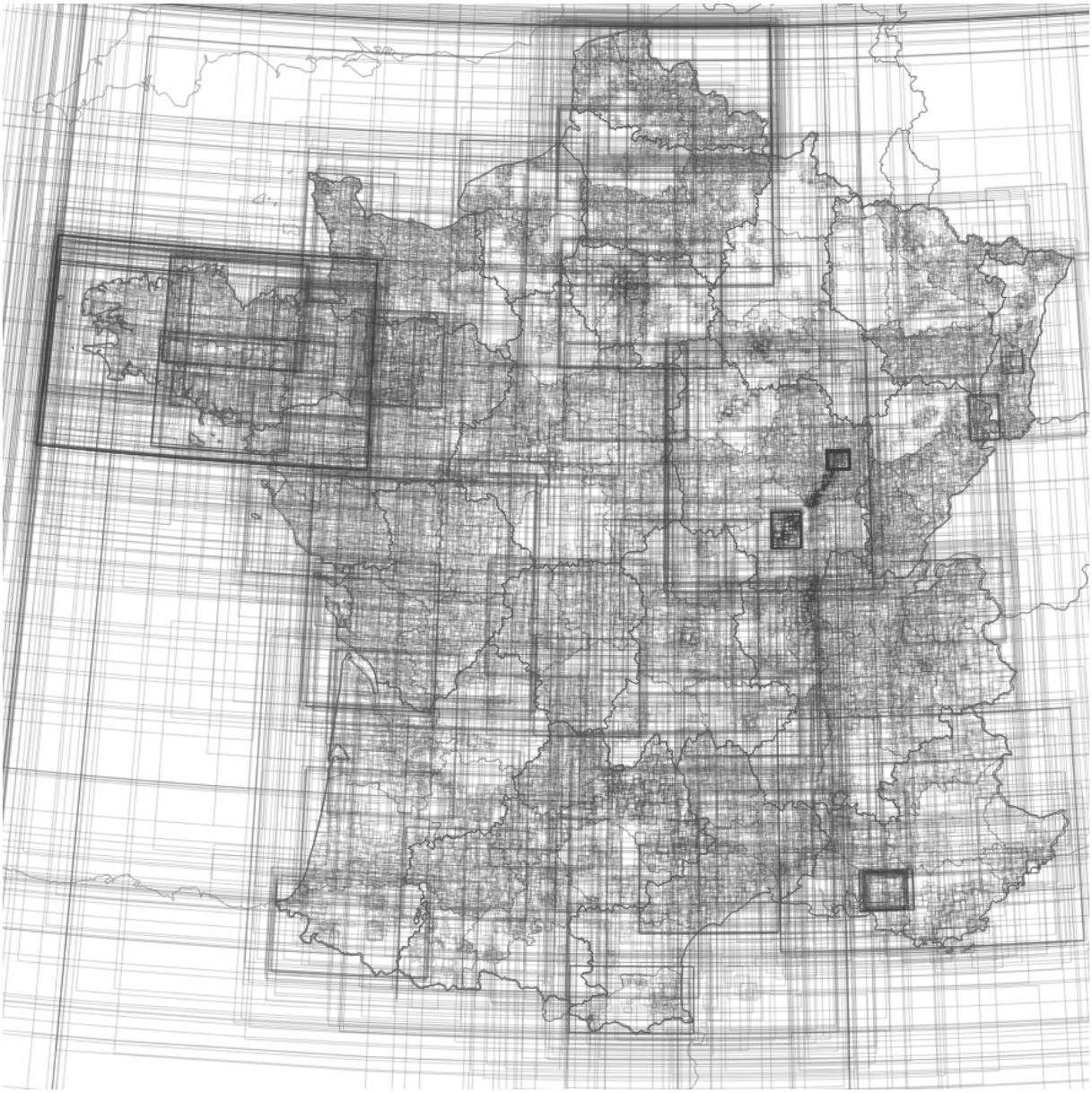

Figure 9. Couche d'emprises « brutes » des métadonnées du géocatalogue (été 2016)

Pour visualiser les recouvrements de ces emprises, une couche de polygones ne se superposant pas est créée, avec en attribut le nombre d'emprises superposées (figure 10).

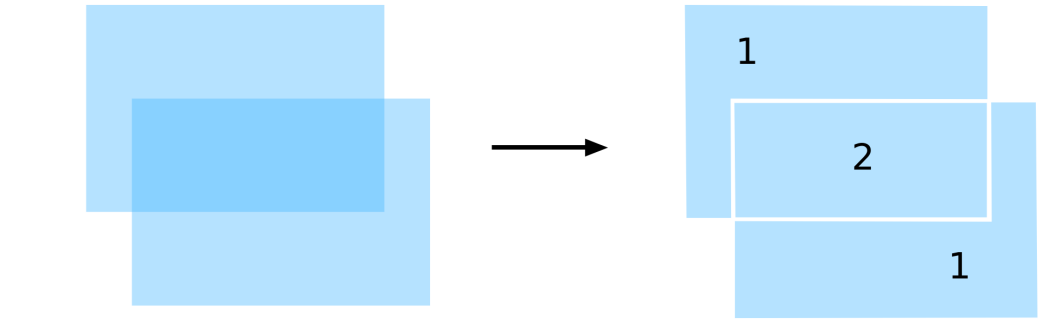

Figure 10. Schéma explicatif du calcul de recouvrements

Cette opération effectuée en SQL en utilisant les opérateurs spatiaux de PostGIS, visualise la *couverture territoriale* de données indexées dans les géocatalogues (figure 11).

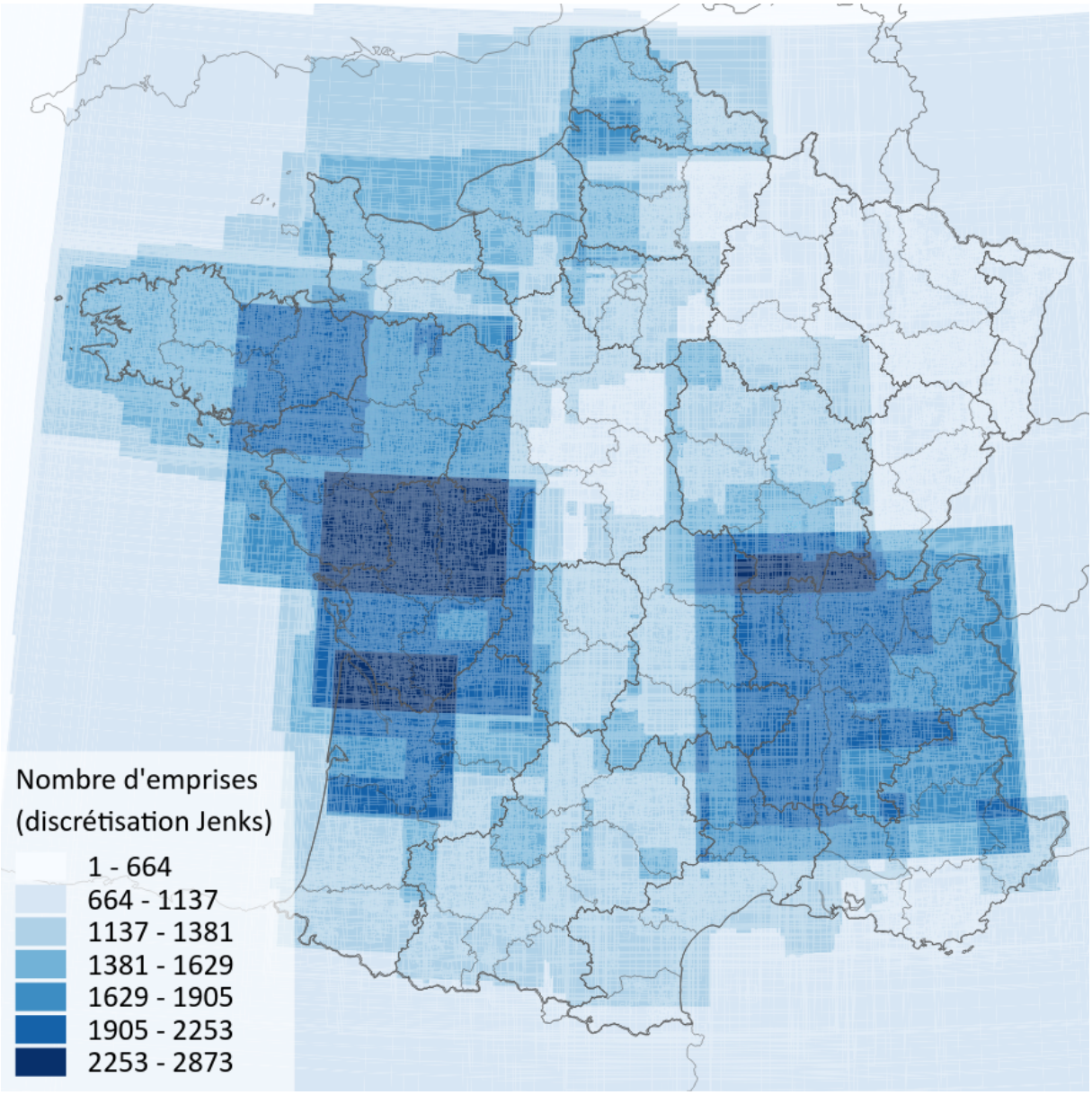

Figure 11. Superposition des emprises des métadonnées du géocatalogue