



Un cadre d'analyse des Infrastructures de Données Géographiques pour interroger la mise en réseaux des acteurs et des outils

M. Noucher¹, F. Gourmelon², A. Laumond³, G. Melançon³, B. Pinault³,
A. Maulpoix², J. Pierson¹, O. Pissot¹, M. Rouan²

1. CNRS, UMR 5319 PASSAGES, Bordeaux Campus

2. CNRS, UMR 6554 LETG-Brest

3. Université de Bordeaux, UMR 5800 LaBRI

GÉOBS :

LES INFRASTRUCTURES DE DONNÉES GÉOGRAPHIQUES

DANS LA GOUVERNANCE INFORMATIONNELLE DE L'ENVIRONNEMENT

PROJET DE RECHERCHE FINANÇÉ PAR LA RÉGION NOUVELLE AQUITAINE (2015-2017)
COORDONNÉ PAR LE CNRS (LABORATOIRES PASSAGES-BORDEAUX ET LETG-BREST)



- **Enjeu scientifique :**
 - Comprendre les stratégies contemporaines de **gouvernance informationnelle de l'environnement**
 - Mener une démarche **dynamique, systémique** et **critique** des dispositifs
- **Objet de recherche :**
 - Les **Infrastructures de Données Géographiques (IDG)**



UMR Passages

BORDEAUX



UMR LETG

BREST



UMR PRODIG

PARIS



UMR LaBRI

BORDEAUX



EA MICA

BORDEAUX

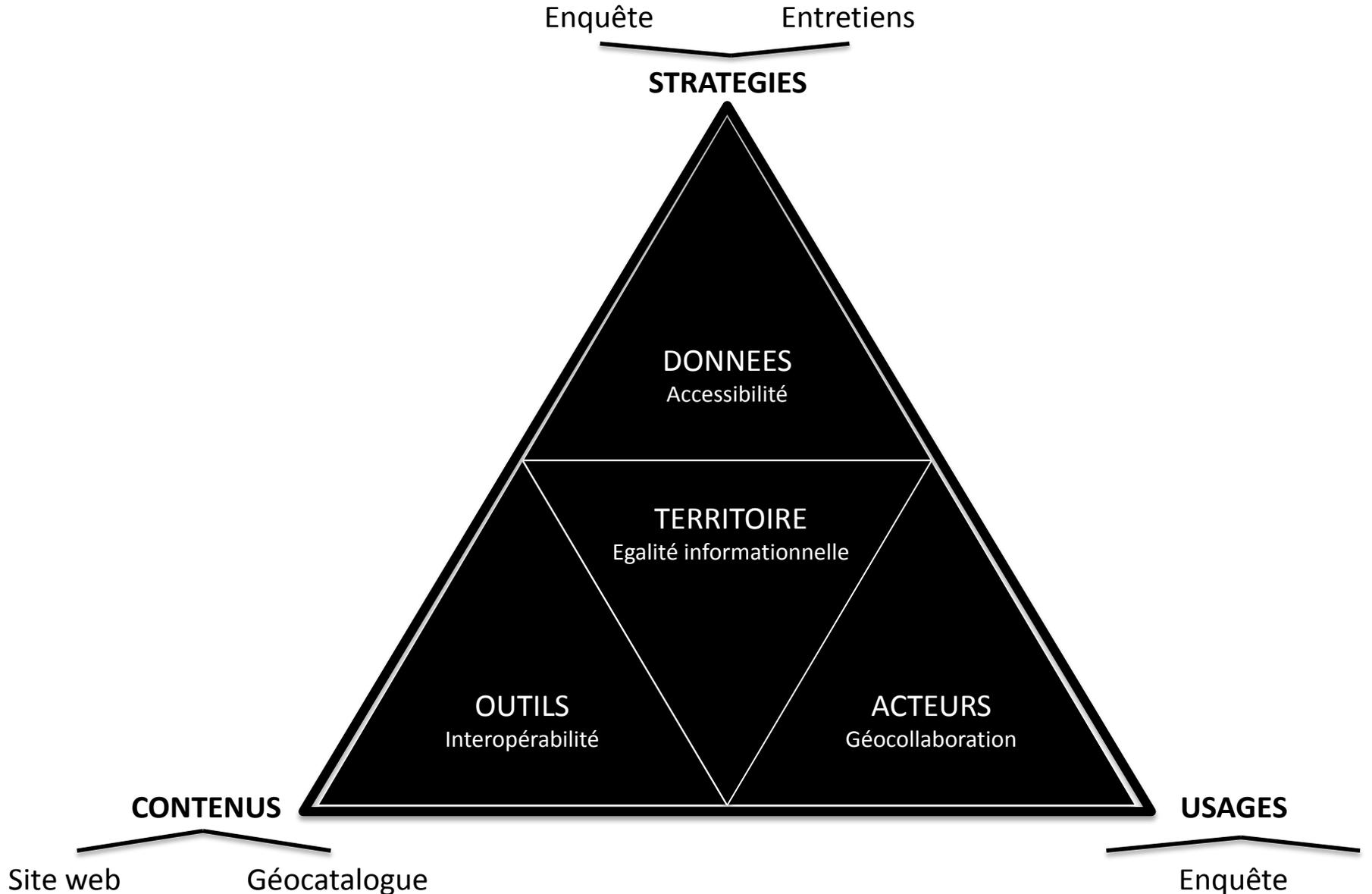
Cadre théorique



Revue Big Data and Society
Special Feature « Critical Data Studies »
<http://bds.sagepub.com/content/critical-data-studies>

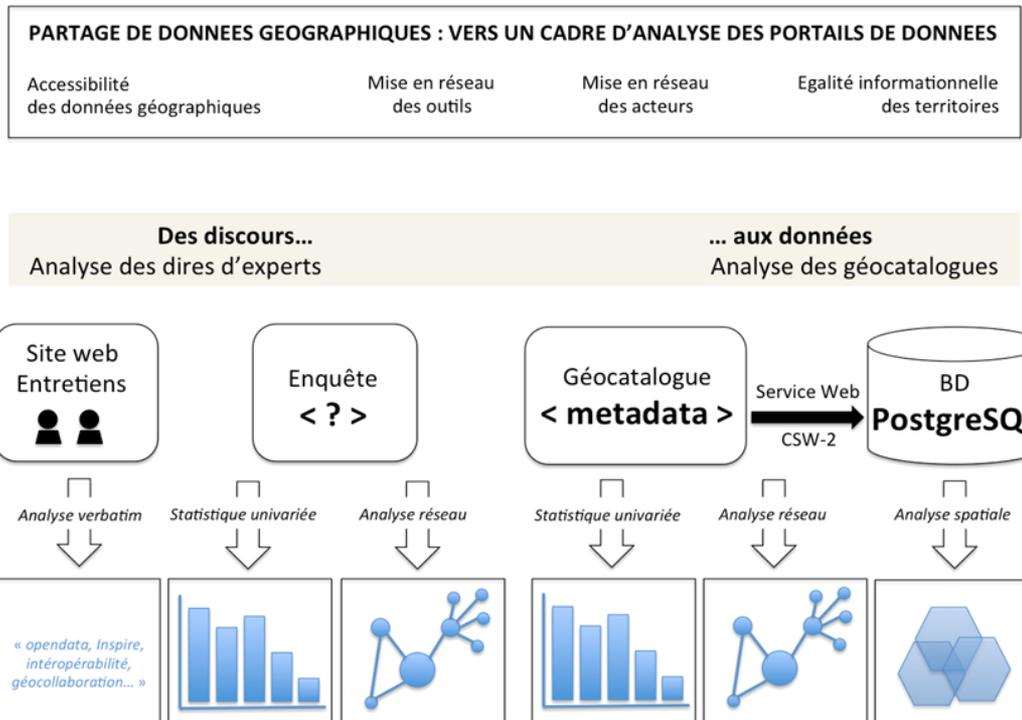
- Les *critical data studies* visent à développer des recherches sur les contextes de production, d'analyse, de diffusion et d'usage des données, en particulier géographiques, qui circulent entre les infrastructures web
- **Action Prospective « Géoweb » du GdR MAGIS** version 2017-2022, dont l'objectif sera de prolonger ces recherches en s'ouvrant sur d'autres objets et d'autres méthodes...

3 axes / 4 dimensions / 45 IDG



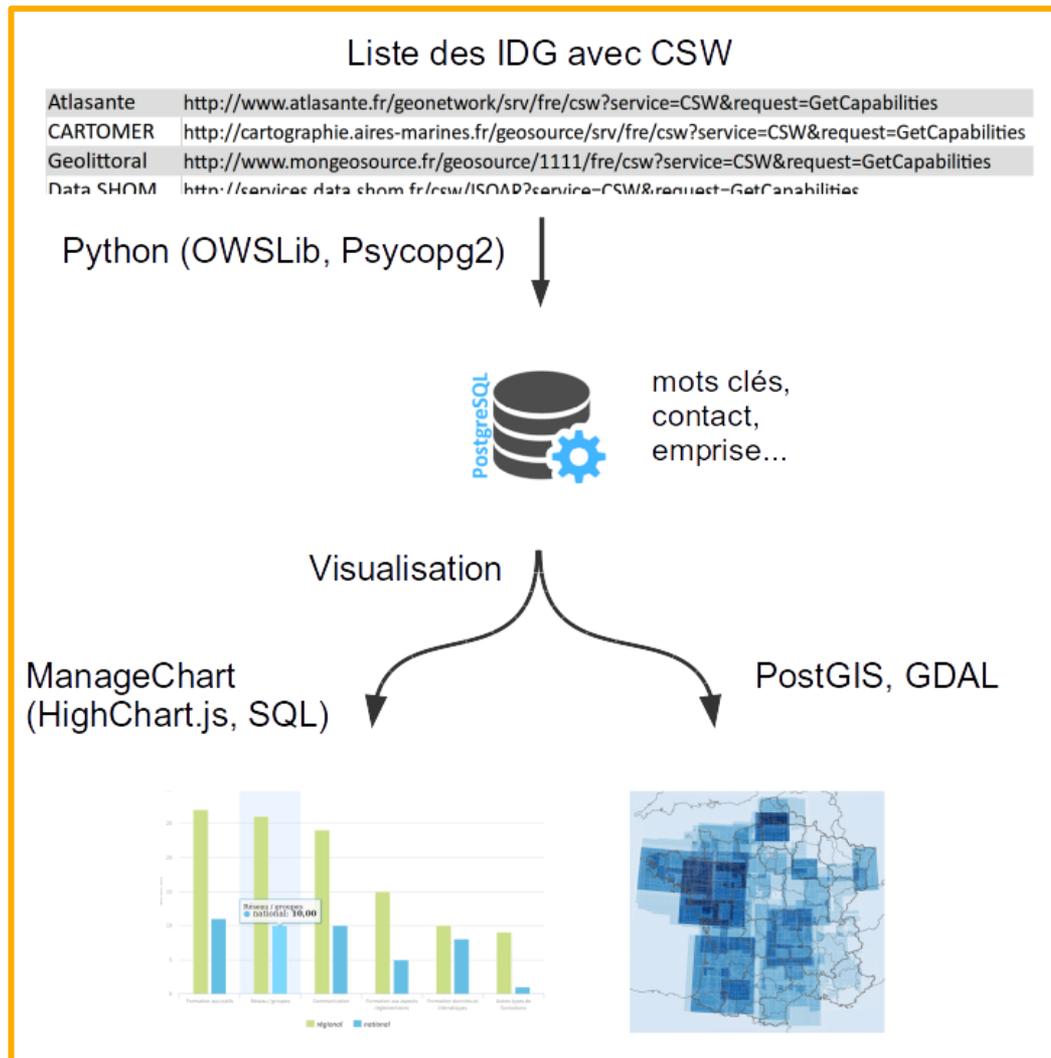
Analyser les 4 dimensions des IDG

- État des lieux des fonctions proposées par les sites
- État des lieux des actions d'animation
- Analyse des géocatalogues
 - Déploiement d'une chaîne de traitement pour extraire, analyser, et visualiser leur contenu



La chaine de traitement CSW Harvester

cf. Atelier EXCES – SAGEO 2016



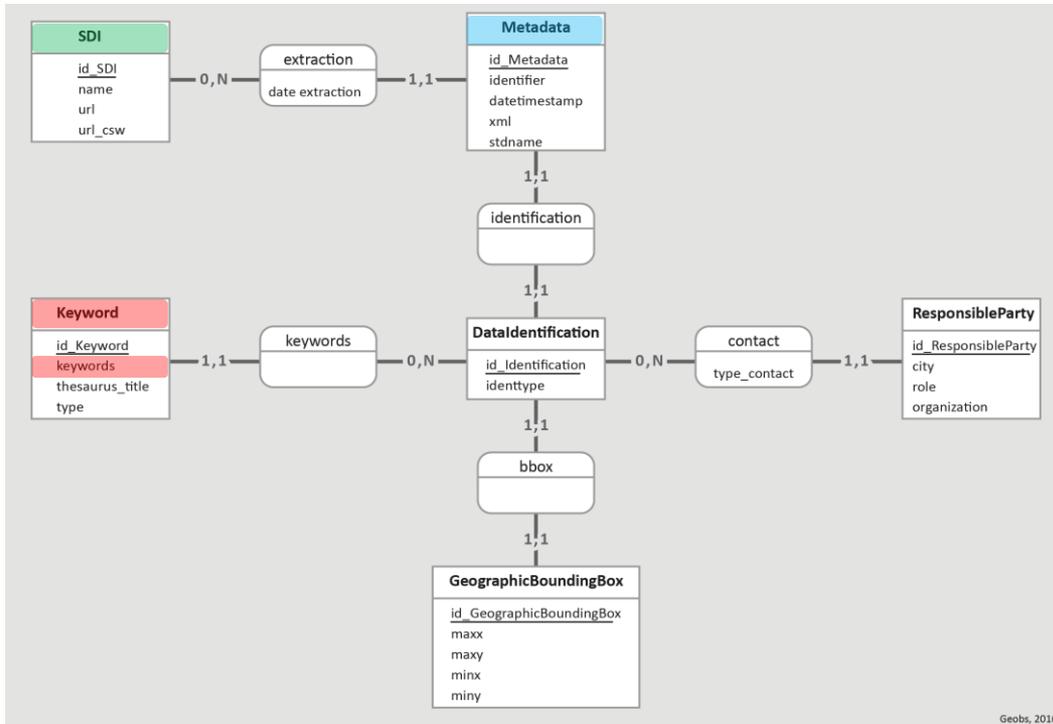
Note méthodologique : <http://www-iuem.univ-brest.fr/pops/attachments/1142>

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

chaîne de traitement : extraction et archivage

extraction informations statistiques, exemple de l'accessibilité des données

BDD CSW-Harvester



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;

données ouvertes

Pattern matching :

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

↳ compter le nombre de **métadonnées** incluant les **mots-clefs** « données ouvertes » ou « open data » et d'en faire le ratio par rapport au nombre total de métadonnées par **IDG**

La chaîne de traitement : analyse et visualisation

extraction informations statistiques, exemple de l'accessibilité des données

ManageChart Bases de données Requetes SQL Graphiques Utilisateurs mathias.rouan@univ-brest.fr

Edition

Nom: geobs - accessibilité d Crédits: GEOBS

Titre: Accessibilité des donn Uri Crédits: http://www-ium.univ-l

Sous-titre: Filtrage des métadonn Légende:

Titre axe-X: Export impression:

Unité axe-X: Export CSV:

Type axe-X: Catégorie Inverser les axes:

Ajouter un axe-Y Enregistrer

Axe-Y n° 1 Titre: Accessibilité (%) Type: Linéaire

Annuler

Ajouter toutes les séries

Sélectionner une requête

Ajouter les séries

Ajouter une série

Ajouter une série

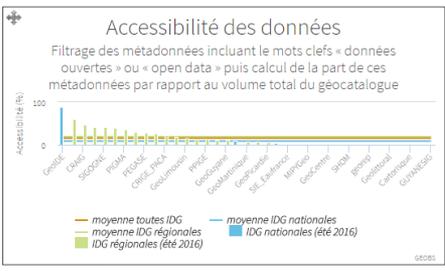
Série n° 5 Annuler

Titre: IDG régionales (été 21) Unité:

Requête: geobs - test Paramètre: régional

Type: Colonne verticale Couleur: #CBDEB8

Marqueur: Style de ligne: Ligne

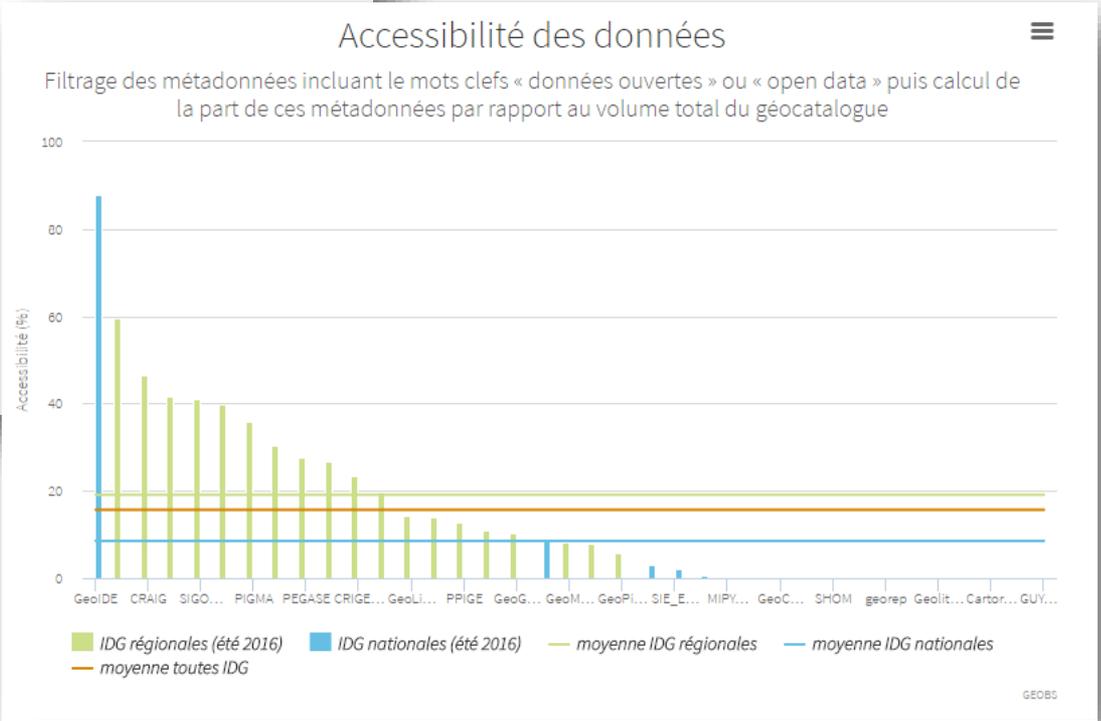


Application web ManageChart

<https://www-ium.univ-brest.fr/wapps/managechart/>

Code source:

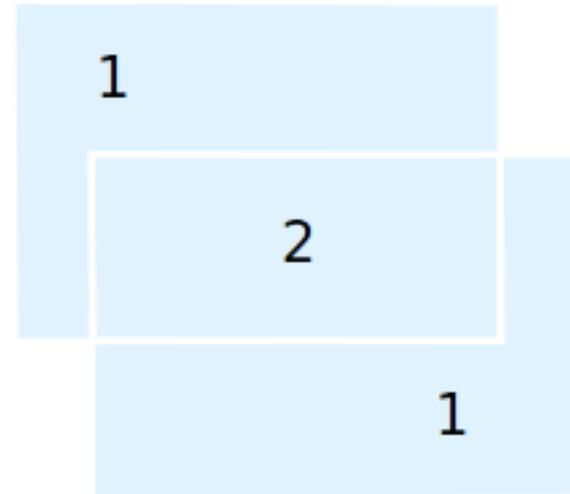
<https://tucuxi.univ-brest.fr/projects/managechart>



La chaîne de traitement : analyse et visualisation

extraction informations spatiales, exemple du recouvrement des emprises

Objectif : couche de rectangles → compter le nombre de recouvrements



2 polygones partiellement superposés

3 polygones sans superposition, avec le nombre de recouvrements en attribut

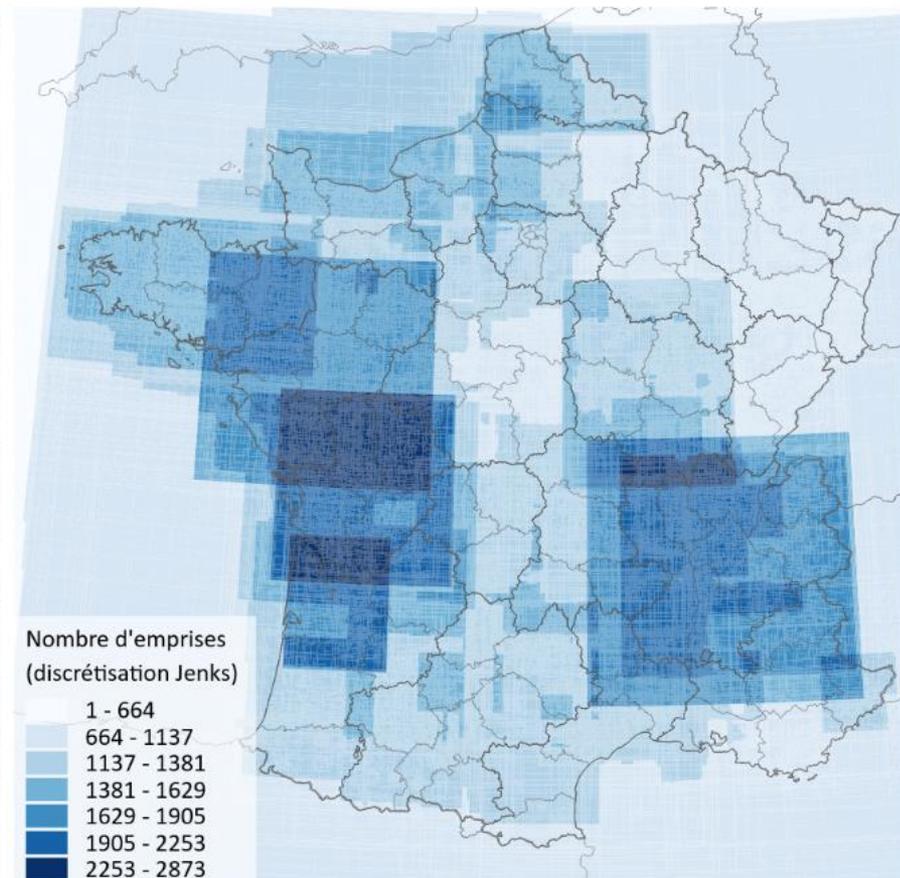
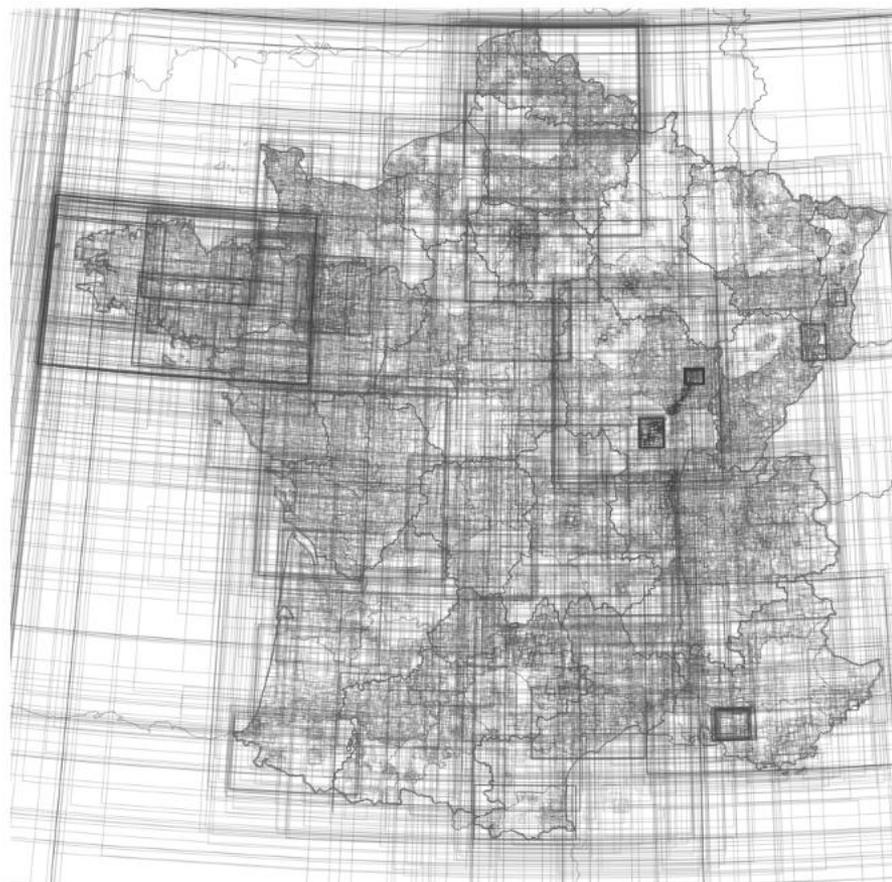
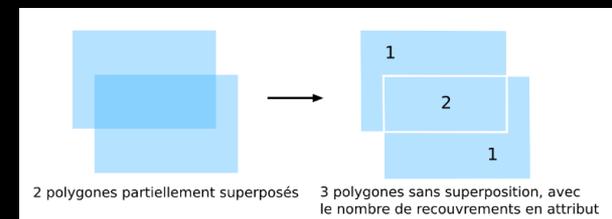
2 solutions testées :

- PostGIS (vecteur)
- Python (rastérisation)



Couverture spatiale

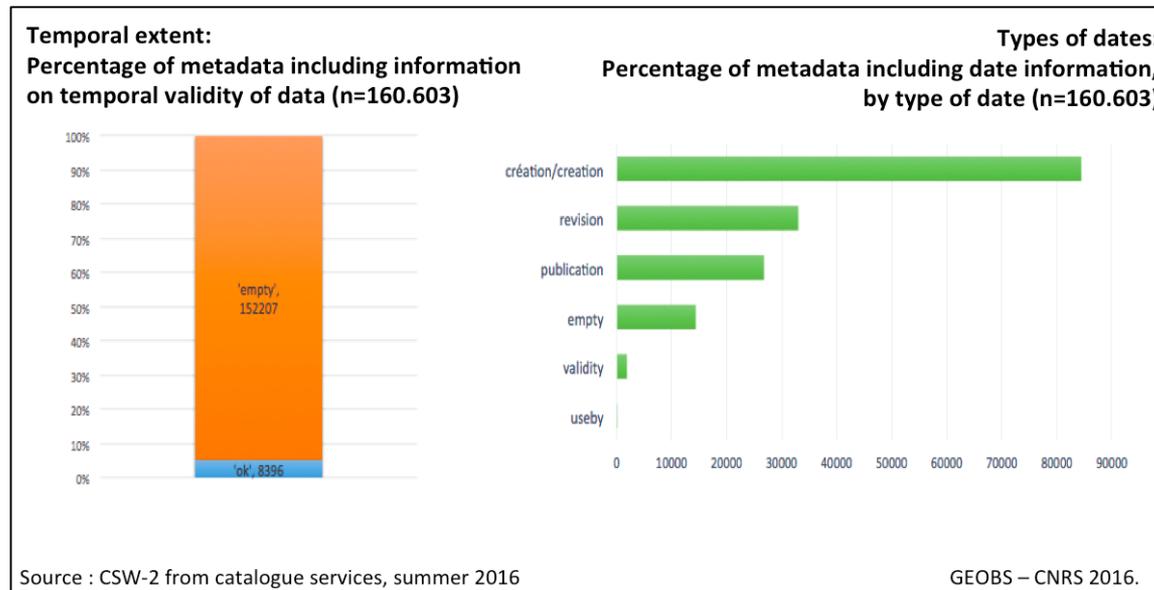
Méthode : extraction / découpage / calcul superposition



Des emprises brutes du géocatalogue (été 2016)... au calcul de leur superposition

Concernant la qualité des métadonnées

- Sur les **45 IDG** étudiées : **37 CSW** opérationnels
→ **160 603 métadonnées** extraites à l'été 2016
- De multiples champs inexploitable :
 - Généalogie, dates, validité temporelle...

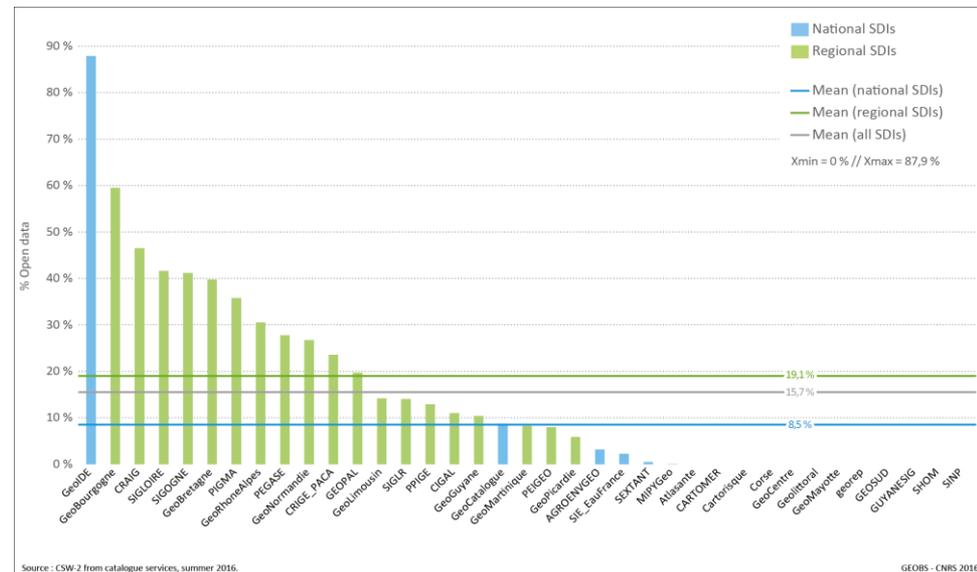
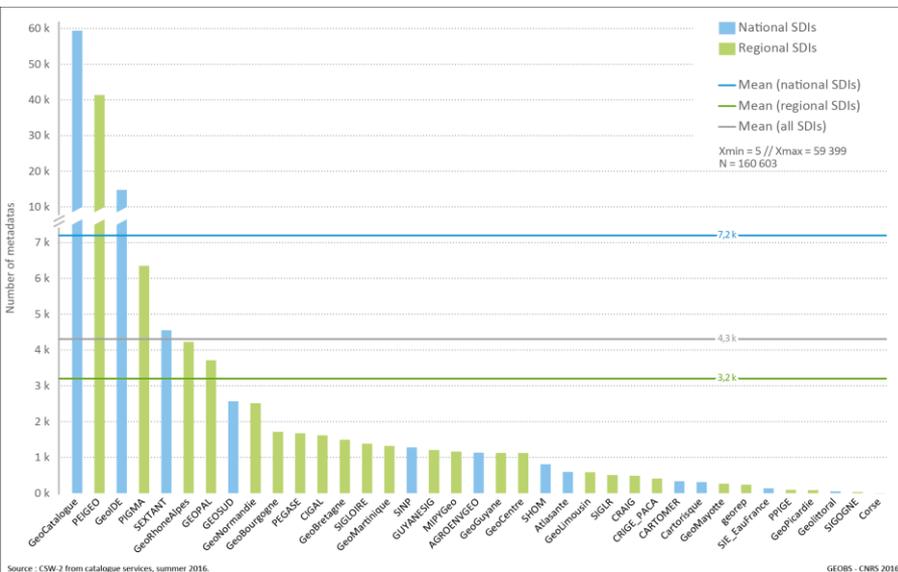


- Mais la possibilité d'interroger la **couverture thématique**, **organisationnelle** et **spatiale** des données indexées dans les IDG

Le contenu des catalogues : quelques résultats

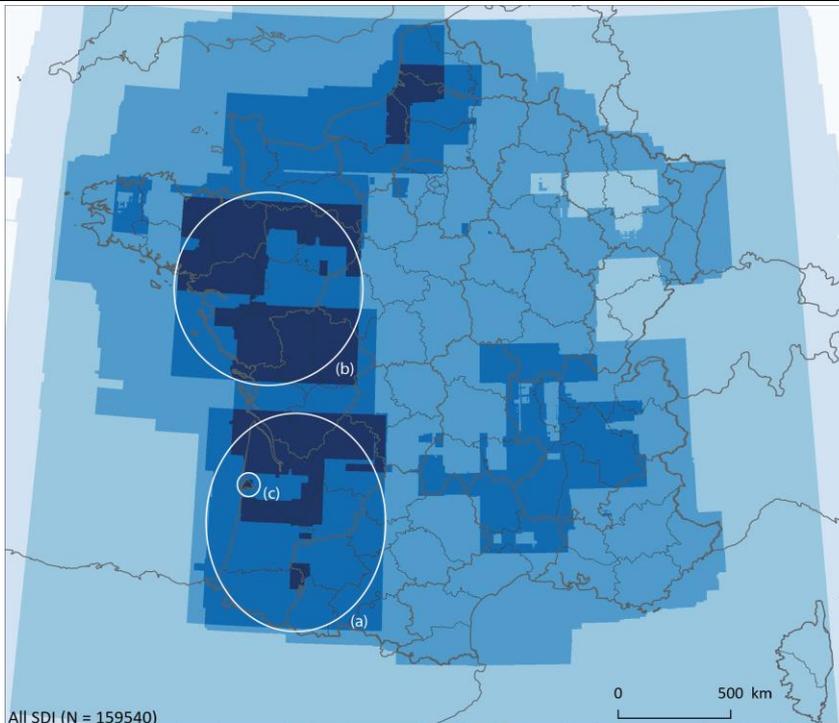
Un volume de métadonnées important...

... mais peu d'informations accessibles

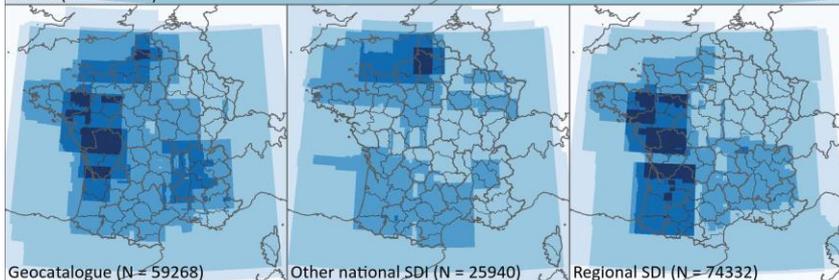


Couverture spatiale

Résultats : des disparités qui interrogent les discours sur l'égalité de traitement informationnel des territoires



All SDI (N = 159540)



Geocatalogue (N = 59268)

Other national SDI (N = 25940)

Regional SDI (N = 74332)

About the data

All data extents are extracted from SDI using CSW. The number of superpositions between these data extents are then calculated.

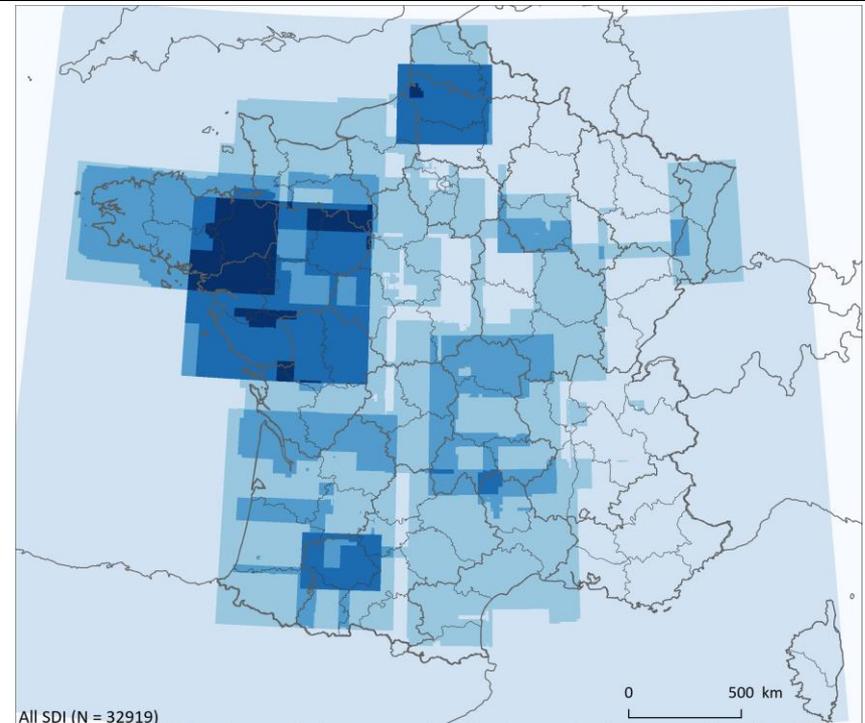
About the discretization

All four maps use the equal interval discretization method : the entire range of data values (max - min) is divided equally into 6 categories.

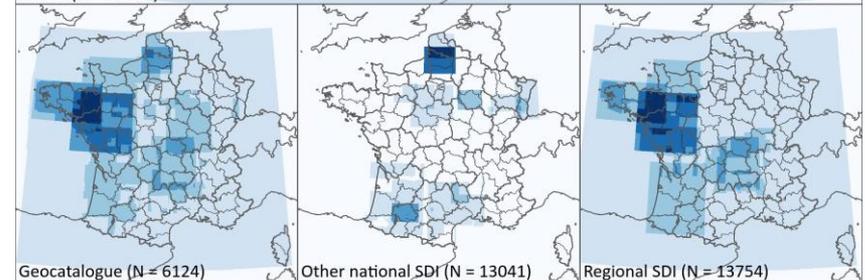
nb of superpositions



$$n = (\max - \min) / 6$$



All SDI (N = 32919)



Geocatalogue (N = 6124)

Other national SDI (N = 13041)

Regional SDI (N = 13754)

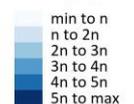
About the data

Data extents from data tagged as "opendata" are extracted from SDI using CSW. The number of superpositions between these data extents are then calculated.

About the discretization

All four maps use the equal interval discretization method : the entire range of data values (max - min) is divided equally into 6 categories.

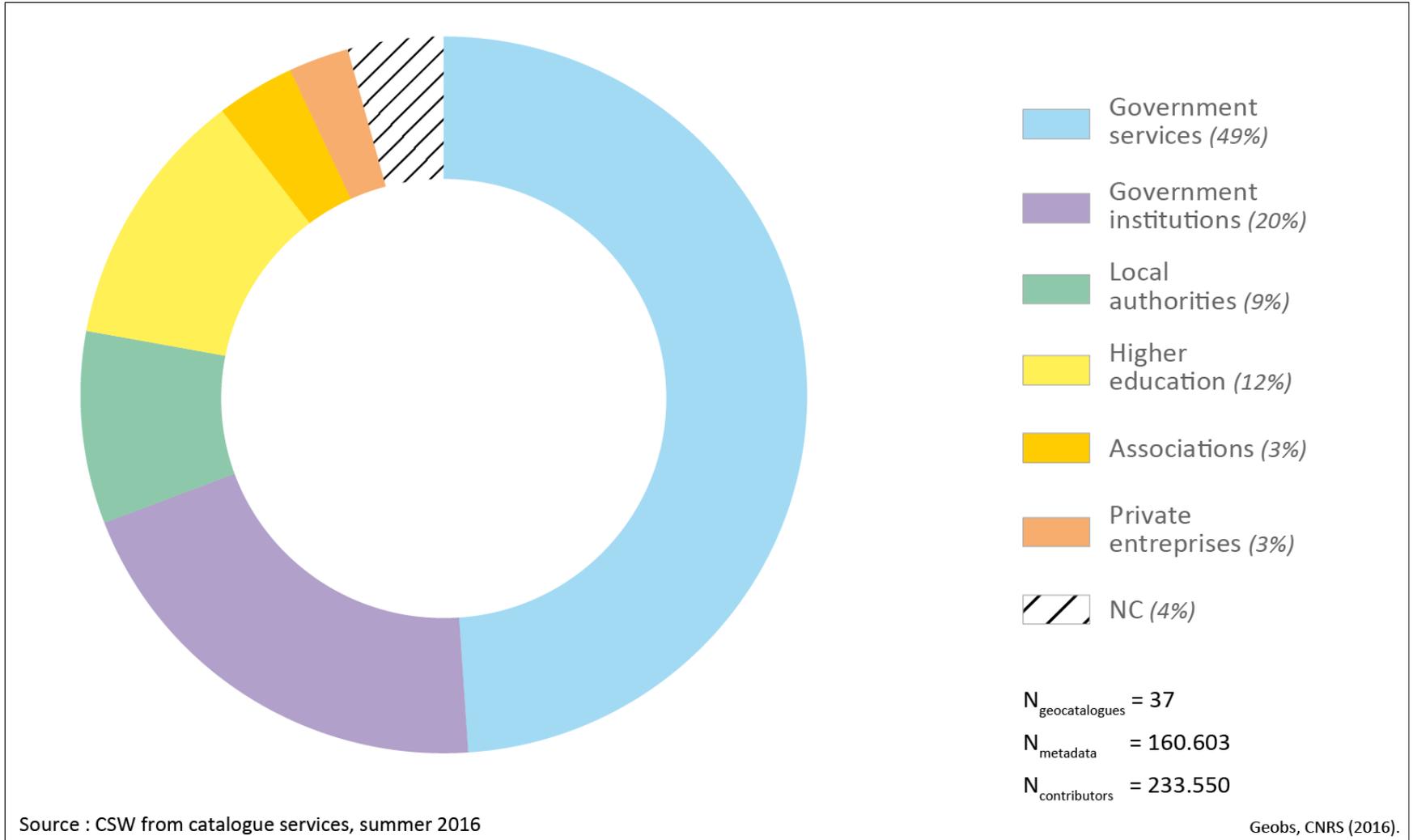
nb of superpositions



$$n = (\max - \min) / 6$$

Couverture organisationnelle

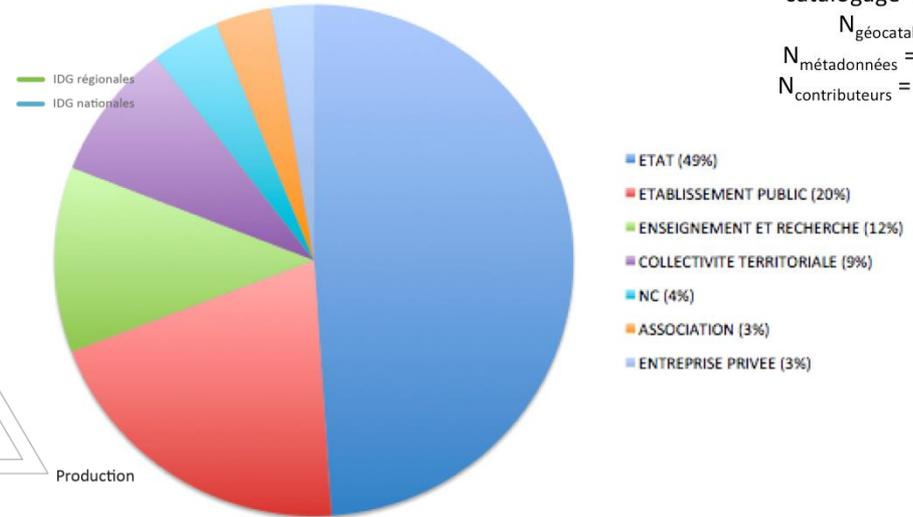
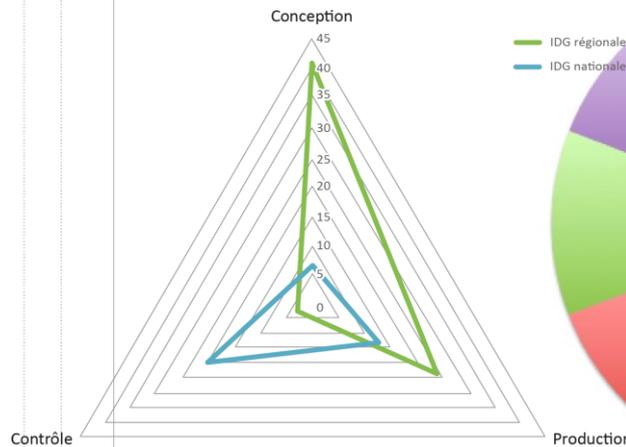
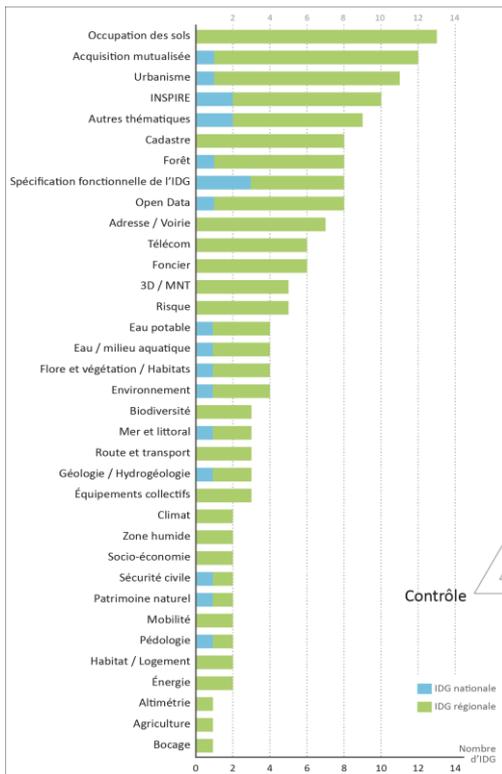
Résultats : 78% des contacts recensés sont issus des autorités publiques



La mise en réseau des acteurs d'après l'enquête et l'examen des sites web

Des fortes dynamiques de géocollaboration...

...entre bénéficiaires assez peu diversifiés



Part des types de contributeurs mentionnés dans les métadonnées des services web de catalogage analysés.

$N_{\text{géocatalogues}} = 37$

$N_{\text{métadonnées}} = 160.603$

$N_{\text{contributeurs}} = 233.550$

Conclusions Perspectives

Enquête Entretiens

STRATEGIES

- analyser les usages
- analyser les changements stratégies / services / usages
- modéliser les IDG

DONNEES
Accessibilité

TERRITOIRE
Egalité informationnelle

OUTILS
Interopérabilité

ACTEURS
Géocollaboration

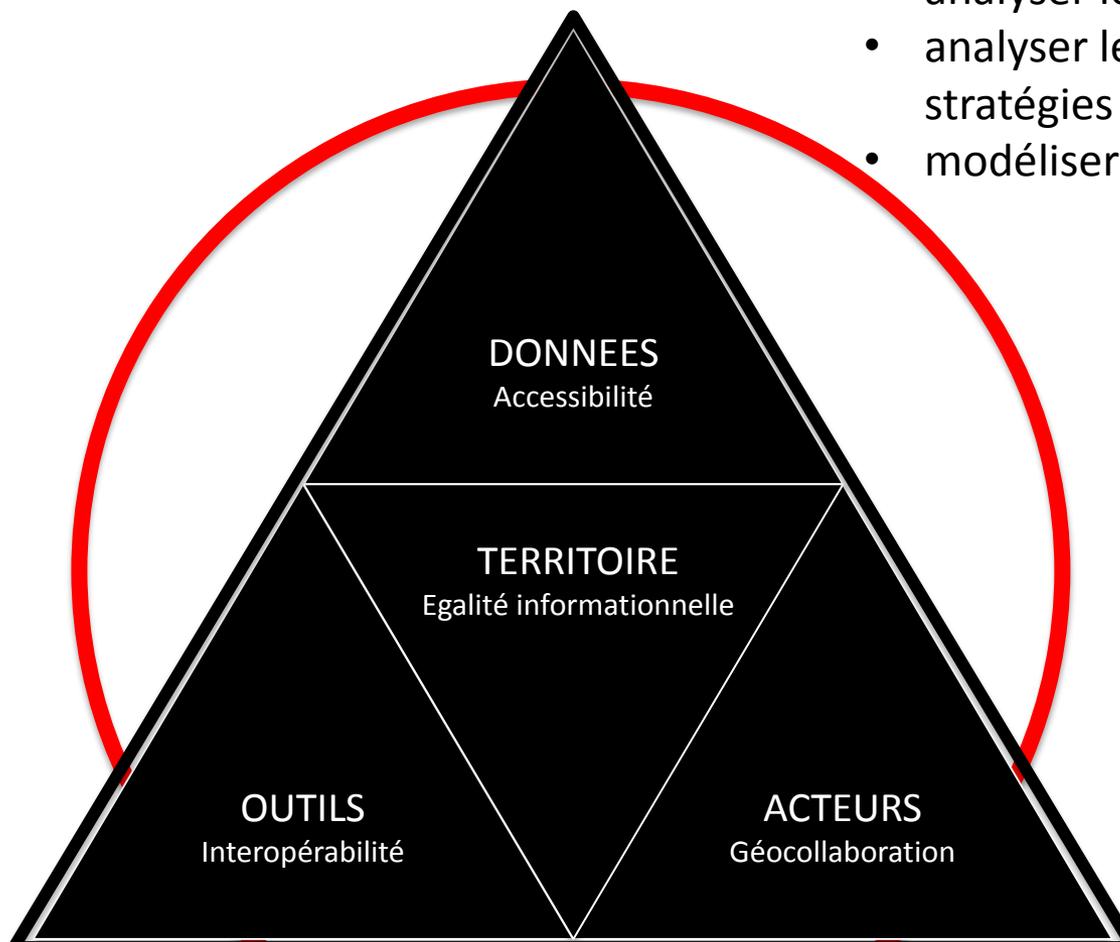
CONTENUS

USAGES

Site web

Géocatalogue

Enquête 2017



Nos canaux de diffusion

- Site web de suivi du projet (documentation, calendrier, intranet équipe) :
<http://www-iuem.univ-brest.fr/pops/projects/geobs>
- Site web de valorisation (visualisations interactives et liens vers ressources) :
<http://geobs.cnrs.fr>
- Dépôt des rapports et articles sur l'archive ouverte institutionnelle **HAL-SHS**
- Dépôt des données documentées sur **data.gouv.fr**
- Dépôt des codes documentés sur **github**



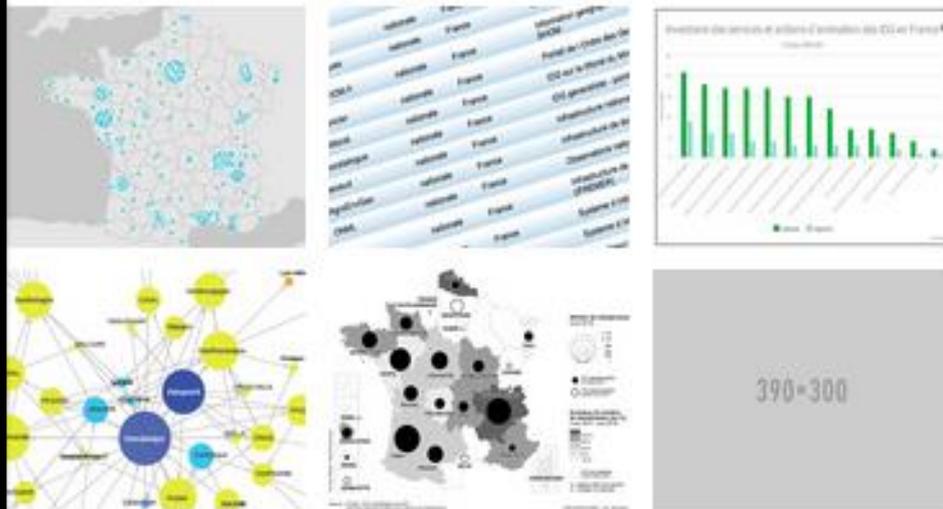
Observer, visualiser et analyser l'information géographique numérique

Ce site rassemble des travaux de recherche qui analysent l'information géographique qui circule sur le web. Il réunit des visualisations qui présentent le contenu, les usages et les flux de données géographiques qui transitent par les plateformes institutionnelles, commerciales ou collaboratives. Il s'inscrit dans l'émergence d'un courant de recherche autour des *critical data studies* qui vise à décrypter les contextes de production, d'analyse, de diffusion et d'usages des données géographiques qui circulent sur les infrastructures web pour en révéler les enjeux socio-politiques.

Les résultats sont présentés sous la forme de (géo-)visualisations qui sont documentées, commentées et pour lesquelles les corpus de données et scripts de traitement sont rendus accessibles.

Accéder au [manifesto](#) du site.

Accéder aux visualisations commentées et aux données et scripts associés.



Accéder aux visualisations commentées et aux données et scripts associés.

