

ANALYSE DES PATRIMOINES NATIONAUX DE DONNÉES GÉOGRAPHIQUES

Comparaison de trois infrastructures nationales de données géographiques (Bolivie, Brésil, France)

Julie PIERSON (ADESS, CNRS Bordeaux)
Matthieu NOUCHER (ADESS, CNRS Bordeaux)
Pierre GAUTREAU (PRODIG, Université Paris 1)
Louca LERCH (Université de Genève)
Olivier PISSOT (ADESS, CNRS Bordeaux)
Aurély JAUTARD (ADESS, CNRS Bordeaux)
Sylvain LESAGE (GeoBolivia)

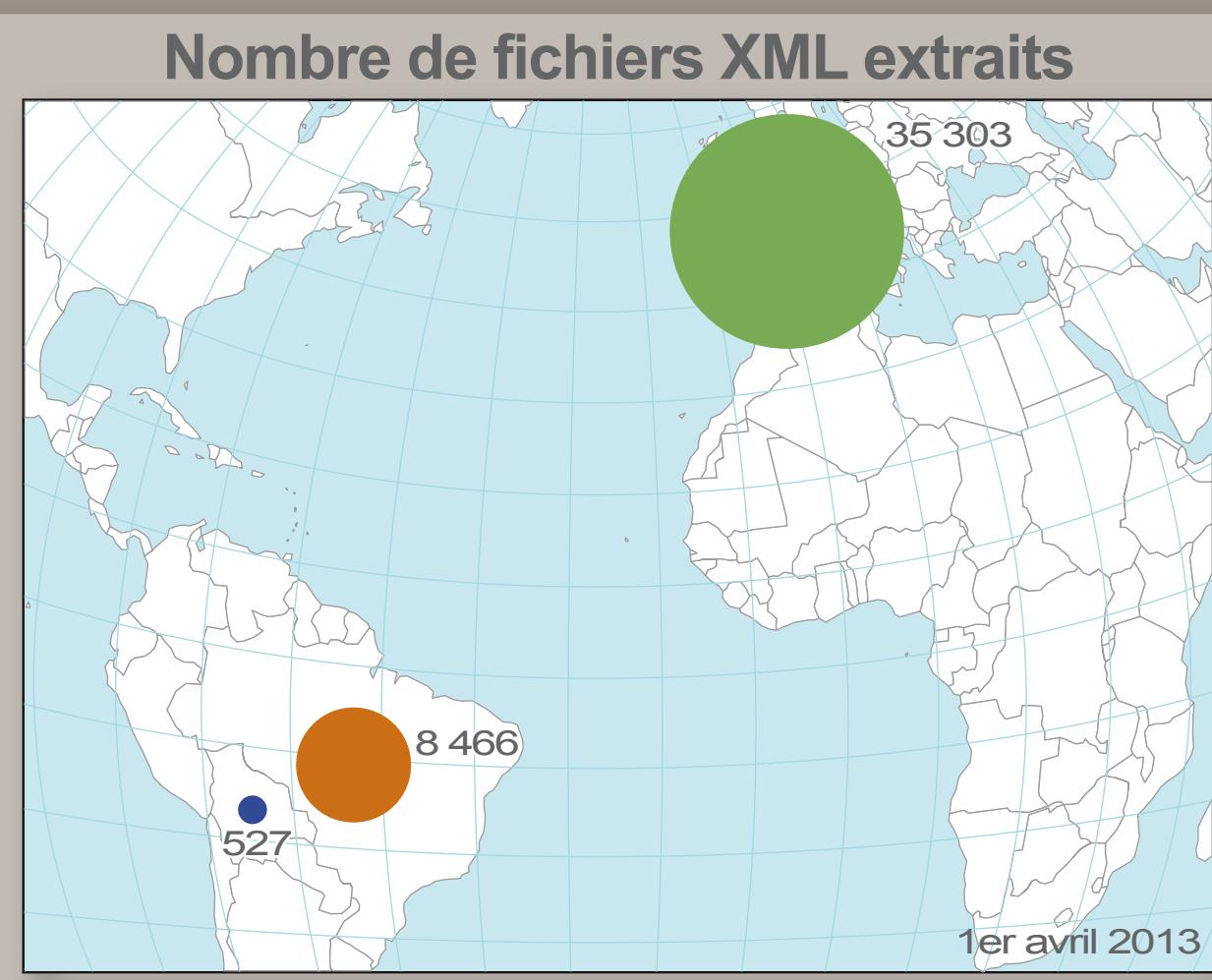


Par le double effet de l'interopérabilité des systèmes et de l'évolution du cadre légal, les **patrimoines de données géographiques institutionnelles** tendent aujourd'hui à être de plus en plus facilement accessibles à travers la mise en place généralisée d'**infrastructures de données géographiques** (IDG). L'analyse comparée du contenu de trois IDG nationales (Bolivie, Brésil, France) permet alors de réinterroger les modalités de formalisation et de diffusion des connaissances sur les territoires.

Pour chacun des **géocatalogues nationaux** étudiés, les **fiches de métadonnées** au format XML ont été extraites au moyen d'un script en python créé pour l'occasion. Les fichiers XML extraits répondent aux normes internationales **ISO 19115** (contenu des métadonnées) et **ISO 19139** (structure des métadonnées).

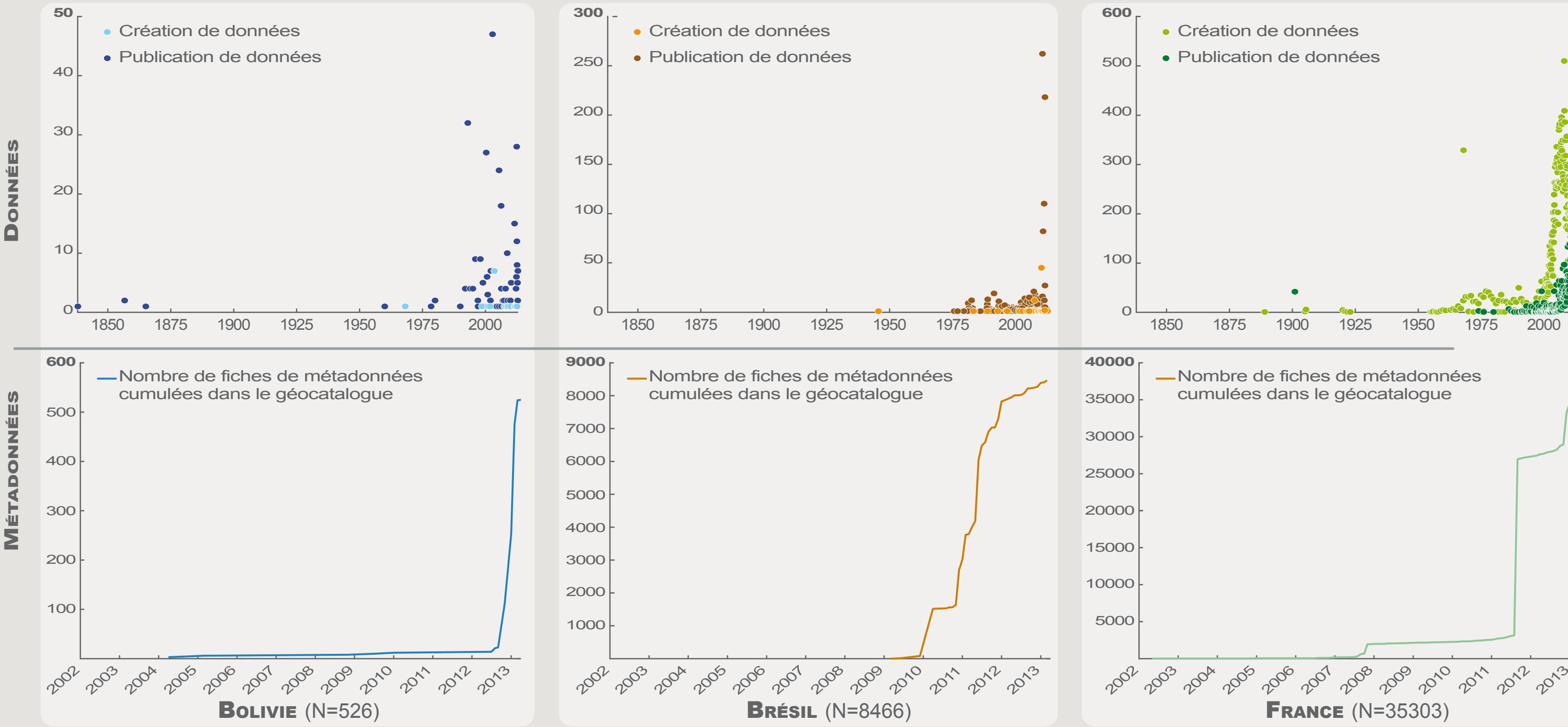
Ils permettent donc une comparaison selon **trois axes d'analyse** :

- **temporel** (à partir des dates de publications des données et métadonnées)
- **thématische** (à partir des classements dans les théâtres)
- **géographique** (à partir de l'emprise spatiale des jeux de données diffusés).



Quand ?

Analyse des dates de constitution des patrimoines de données



Les rythmes différents d'alimentation des catalogues peuvent s'expliquer (en particulier pour les dates de publication des métadonnées), par l'évolution du cadre réglementaire (la publication du règlement INSPIRE le 3 déc. 2008 marque le début d'une montée en charge du géocatalogue français) ou par des logiques organisationnelles (mise en place d'ateliers de saisie de métadonnées en Bolivie depuis mi-2012).

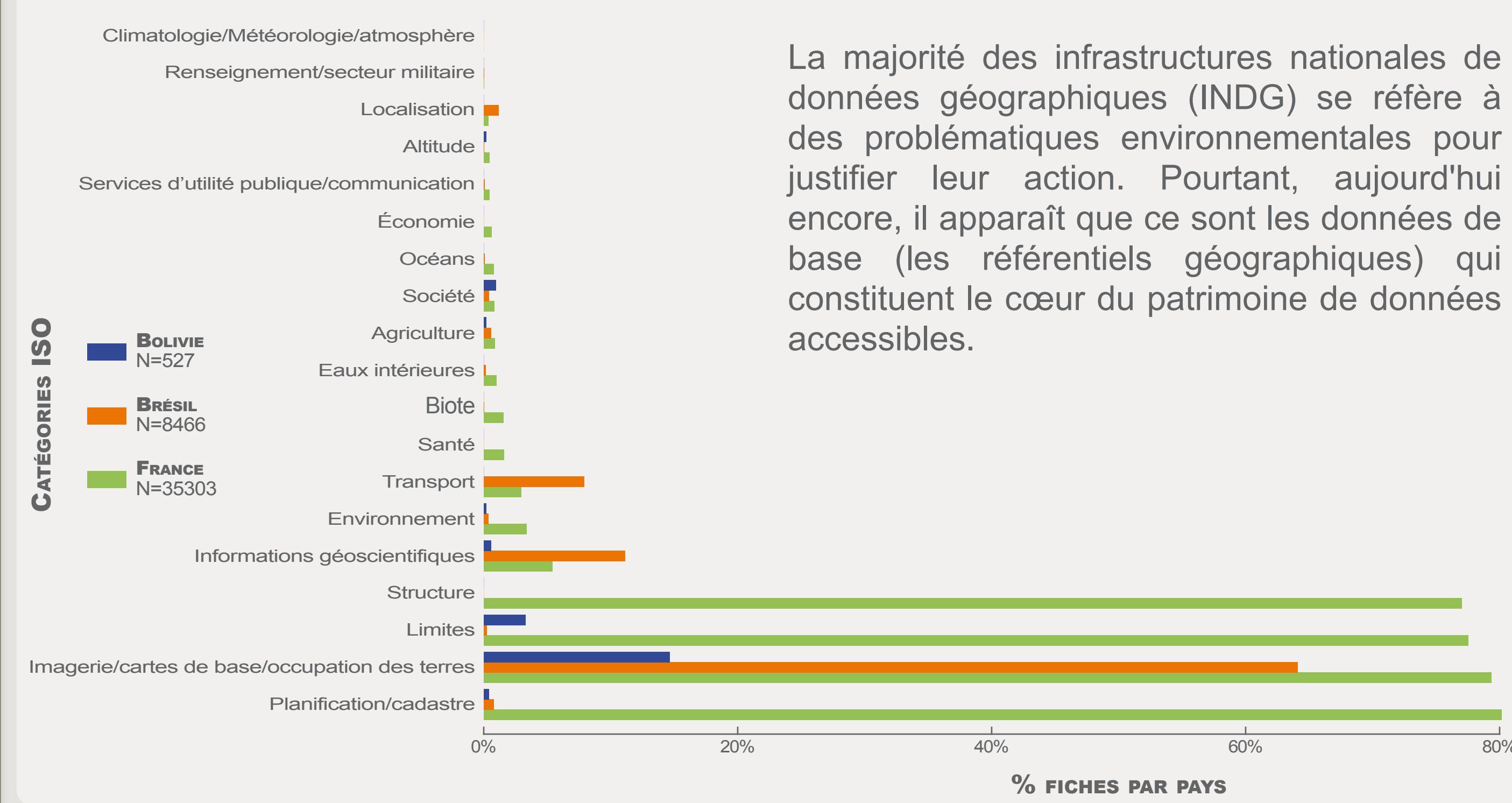
Focus méthodologique : La datation de la constitution des patrimoines de données peut se faire par l'analyse de trois balises XML relatives à la date de création et/ou la date de révision et/ou la date de publication de la donnée. Cependant, la saisie de ces trois types de dates est très inégale d'une fiche de métadonnées à l'autre et d'un catalogue à l'autre. Dans ces conditions il devient délicat de ne sélectionner qu'un champ de métadonnées pour comparer les trois catalogues.

Part des métadonnées renseignée par type de "date"

	Bolivie	Brésil	France
Avec date de création (%)	4	54	90
Avec date de publication (%)	71	33	26
Avec date de révision (%)	8	1	20

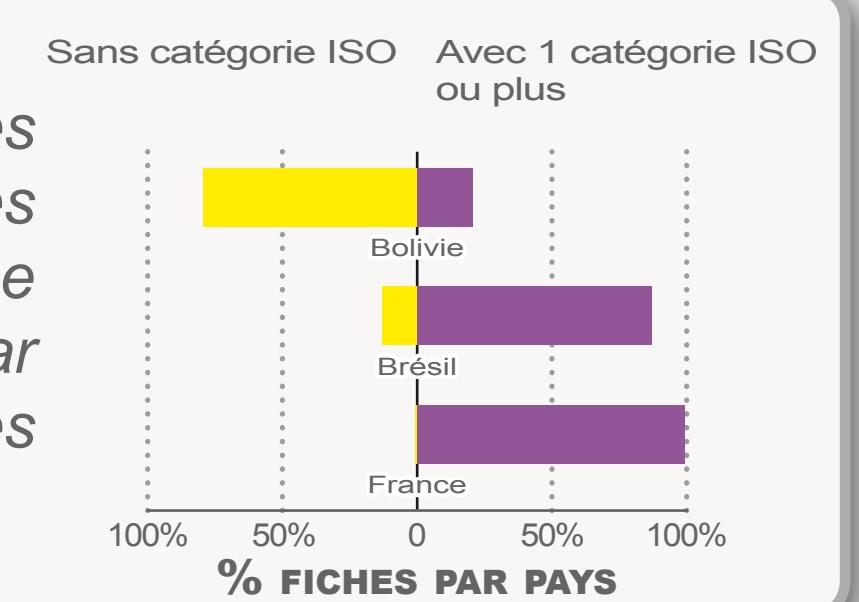
Quoi ?

Analyse des thématiques des données diffusées



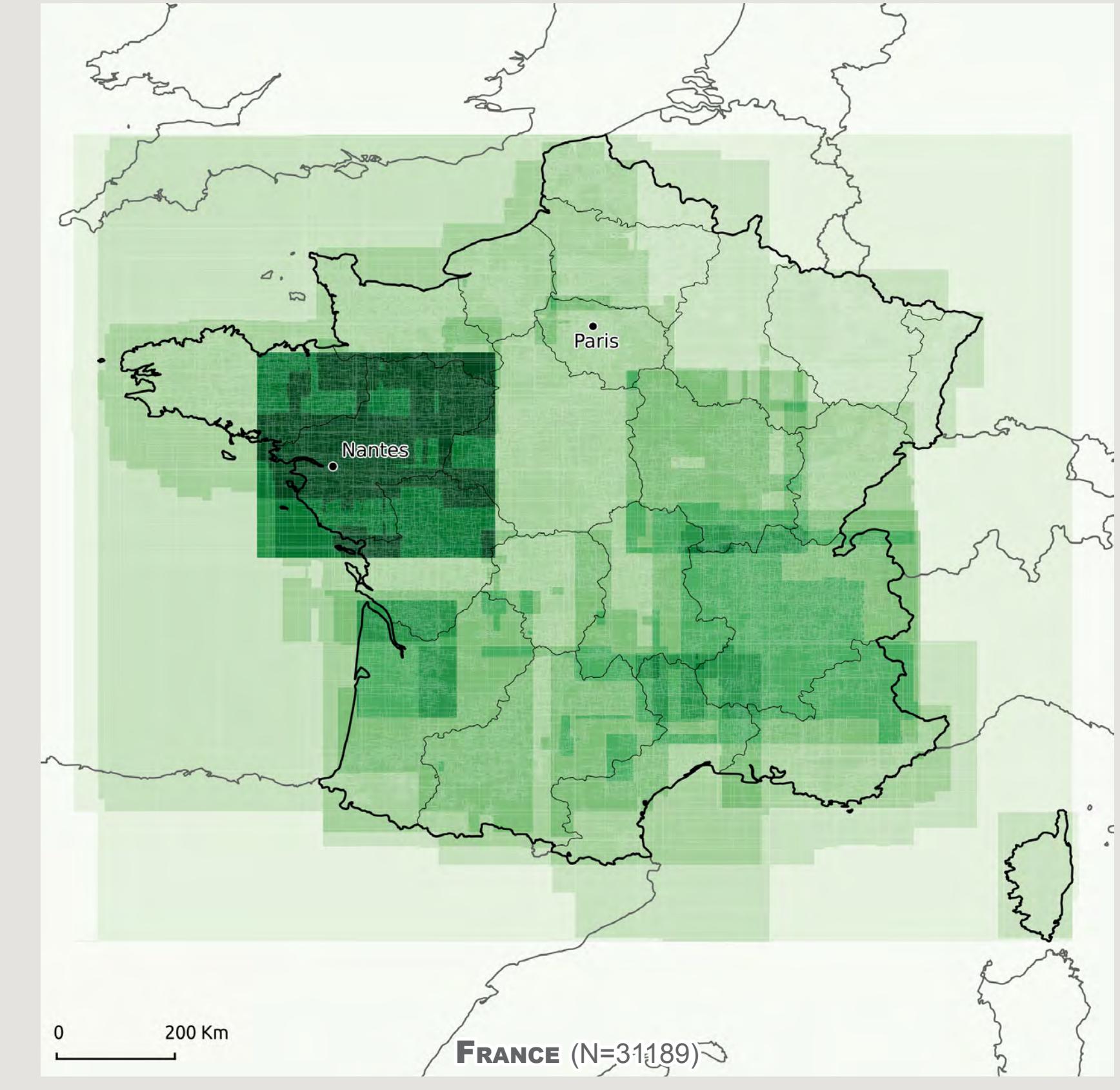
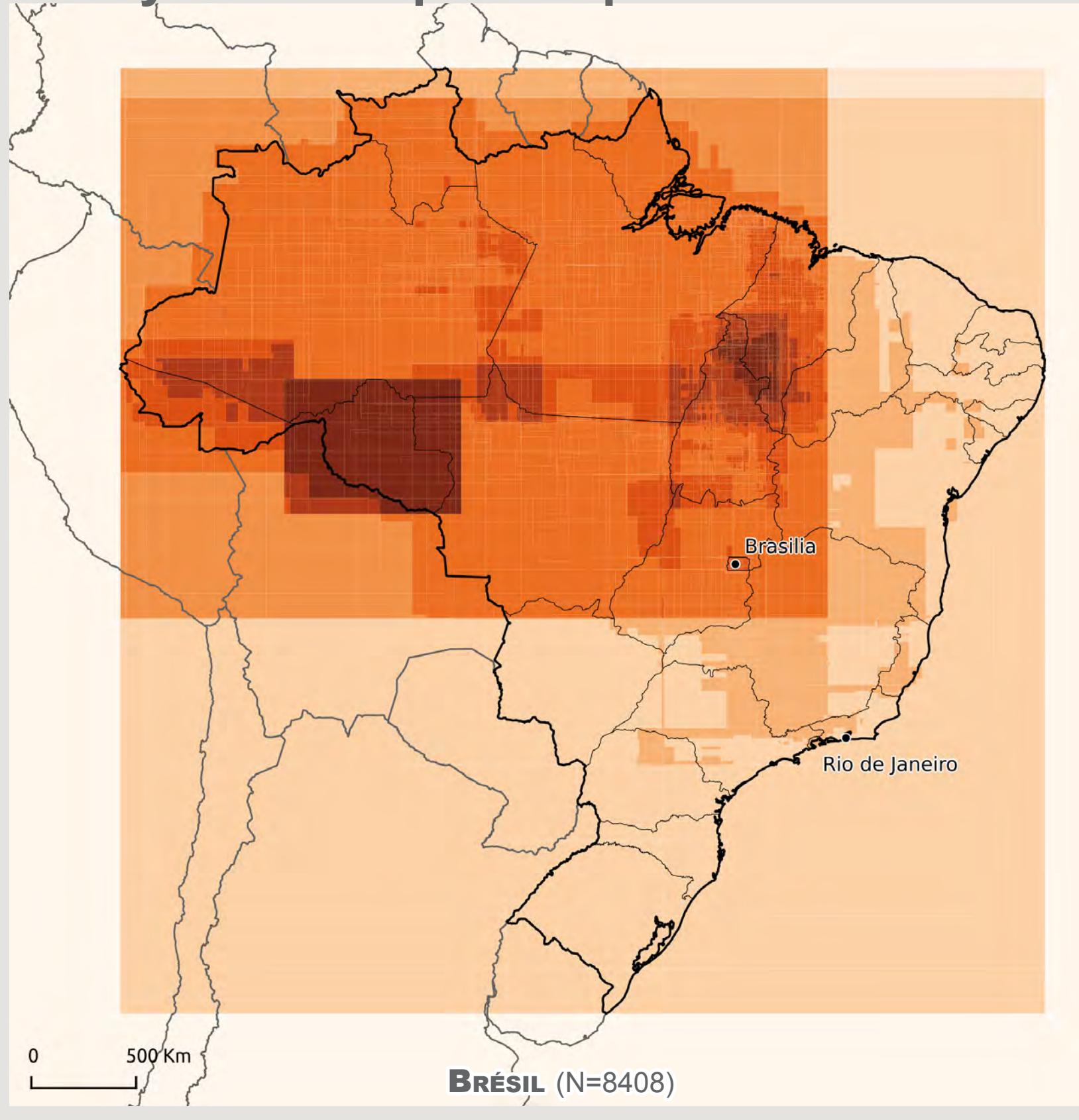
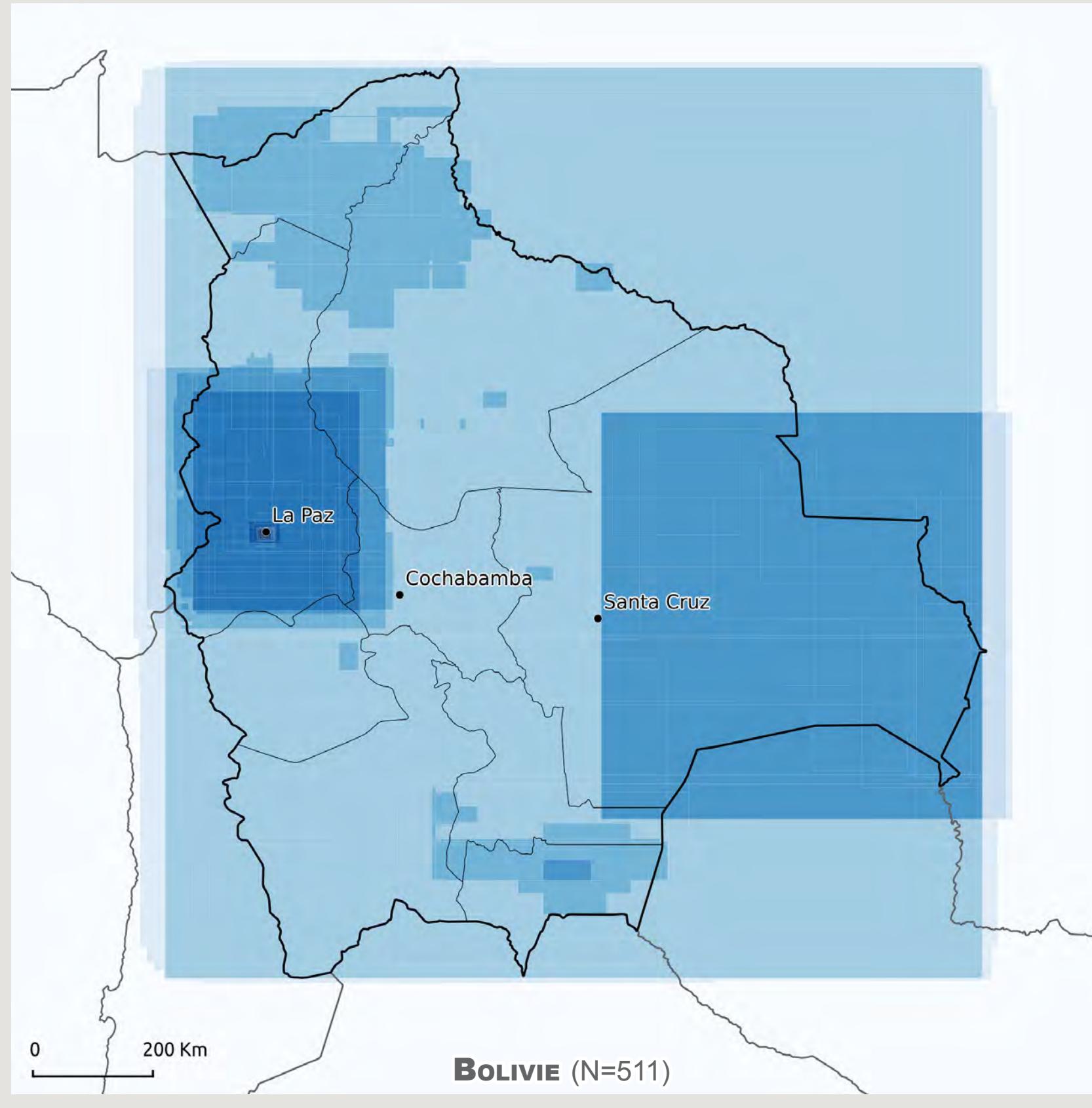
La majorité des infrastructures nationales de données géographiques (INDG) se réfère à des problématiques environnementales pour justifier leur action. Pourtant, aujourd'hui encore, il apparaît que ce sont les données de base (les référentiels géographiques) qui constituent le cœur du patrimoine de données accessibles.

Focus méthodologique : La comparaison des catégories de données s'est révélée complexe dans la mesure où les trois catalogues ne respectent que partiellement la norme ISO19115. Ainsi le théâtre ISO qui permet de décrire par des mots-clés normalisés les données est mobilisé de façon très inégale.



Où ?

Analyse des emprises spatiales des données



Cette spatialisation de l'information géographique révèle des répartitions hétérogènes de l'information et des discontinuités spatiales qui ne sont pas le seul fait de disparités démographiques. En Bolivie par exemple, l'aire urbaine de Cochabamba apparaît comme beaucoup moins couverte que des secteurs quasi déserts comme le Pando au nord ou le Chaco au sud. Au Brésil, l'INDG, portée par le Ministère de l'Environnement se focalise davantage sur le massif amazonien. En France, on distingue nettement l'importance de l'échelon régional avec des configurations différenciées, liées notamment à la présence ou non d'IDG régionales moissonnées au niveau national (à l'image de GEOPAL en Pays de la Loire).

Focus méthodologique : L'extraction des 4 coordonnées des rectangles d'emprise (un même jeu de données peut posséder plusieurs emprises si la couverture est discontinue) est réalisée à partir d'un script en XQuery qui permet de récupérer des champs d'un catalogue, d'en extraire les balises XML correspondantes aux emprises (<Ex_GeographicBoundingBox>) pour générer ensuite une couche SIG en format GeoJson, traitée dans PostGIS.

Les extractions et analyses des métadonnées de trois **géocatalogues nationaux** mettent en évidence l'intérêt d'une analyse approfondie des **infrastructures de données géographiques** :

- les IDG disposent d'une « **épaisseur historique** » qui commence à être suffisamment conséquente pour offrir un **regard dynamique sur la constitution et la circulation de l'information géographique institutionnelle** ;
- l'intérêt de ces premières analyses est aussi de montrer les **difficultés méthodologiques** associées à l'exploitation des métadonnées ;
- les **écart aux normes** internationales se révèlent alors être à la fois une difficulté technique pour engager la comparaison de catalogues différents et une opportunité pour analyser les interprétations divergentes et adaptations locales qui révèlent la **variété des dynamiques organisationnelles et territoriales** sous-jacentes à ces dispositifs socio-techniques.