

## COMPTE-RENDU DE REUNION

REF : SPP/22.0438

DATE : 29/03/2022 – Lieu : virtuel

**Objet :** Réunion du GT Métadonnées – Catalogues et moissonnage

**Ordre du jour :**

Retex Géo2France sur la problématique de moissonnage de catalogues open data à partir d'un catalogue GeoNetwork (Benjamin Chartier)

Retex idéobfc sur les différents moissonnages entrepris (Jerome Boutet)

Présentation autour des évolutions du Géocatalogue (Gregory Delobelle)

Divers

Commission Règle et Qualité

Prochaine réunion

**Principales conclusions :**

Cf. Ci-dessous

**Prochaine réunion :**


La prochaine réunion sera dédiée au mapping INSPIRE/DCAT.

La date est à fixer.

**Liste de diffusion :**

Participants – Organisme / Service		Personnes à informer – Organisme / Service	
Adeline Souf	Shom	Participants et :	
Alix Marc	Ville de Montpellier	Amandine Thomas	Ifremer
Arnauld Gallais	CEREMA	Berenice Lequesne	Shom
Benjamin Chartier	Bchartier	Carlo Mouzannar	Isogéo
Benoit David	CGDD/SRI/ecolab	Dimitri Meunier	OIEAU
Clement Jaquemet	CGDD/SRI/ecolab	Erwann Quimbert	Ifremer
David Viglietti	OIEAU	Guillaume Grech	MNHN
Fanny Lecuy	SHOM	Laurène Debray	OIEAU
Grégory Delobelle	BRGM		Rennes Métropole
Guillaume Ryckelynck	Région Grand Est	Maël Reboux	
Jerome Boutet	IDÉO BFC	Magali Pessidou	DGFIP
Julien Meillon	Ifremer	Noël Cellarier	DGALN
Léo Darengosse	Isogéo	Sébastien Léger	DGFIP
Leslie Lemaire	SG/SNUM/UNI/DRC	Sébastien Dias	Gipatgeri
Maria Tomanov	Région Auvergne Rhone Alpes	Thierry Vilmus	BRGM
Marie Lambois	IGN	Vincent Gomand	DGFIP
Marion Lacroix	IGN	Yvan Le Bras	MNHN
Mathieu Becker	Isogéo		
Robert Rivière	Service du Numérique du MTE		

Thomas Gratier	Etalab	
----------------	--------	--

Date	Visa	Nom	Service
05/04/2022		Marie Lambois	IGN

## Décisions et minutes

### Retex Géo2France sur la problématique de moissonnage de catalogues open data à partir d'un catalogue GeoNetwork (Benjamin Chartier)

Les travaux présentés ont été réalisés fin 2021 à la région Hauts de France, en partenariat avec Neogeo et Camp2camp.

Géo2France est une plateforme régionale partenariale. Elle offre un catalogue mais les partenaires peuvent aussi utiliser leur propre catalogue. Elle est basée sur GeoOrchestra, qui contient Geonetwork et CKAN.

La volonté de ces travaux était de simplifier l'accès aux données, d'être capable de moissonner les catalogues open data des partenaires locaux (types de catalogues très variés).

Un premier axe de travail a été d'étudier le vocabulaire DCAT. Il existe plusieurs déclinaisons du vocabulaire DCAT :

- statDCAT AP : embraque également un catalogue d'attribut
- GeoDCAT-AP (à l'origine basée sur DCATv1, celle basée sur DCATv2 ici <https://semiceu.github.io/GeoDCAT-AP/releases/#introduction>)

Des tests ont également été menés pour voir quelles implémentations étaient réalisées :

- ArcGIS online DCAT en JSON-LD
- OpenDatasoft JSON (pas DCAT) socle en anglais, extensions en français..)
- UData DCAT (minimal) ou avec l'API REST, du JSON plus complet

En conclusion, il a été constaté que souvent il y a moins d'information dans les flux de métadonnées standards (par rapport à ceux qui utilisent un modèle propriétaire). Parfois le moissonnage est compliqué car nécessite plusieurs requêtes pour accéder à la totalité de l'information. Le choix a donc été fait de ne pas utiliser DCAT comme format pivot.

Une liste a donc été élaborée pour déterminer quels éléments de métadonnées sont utiles à l'utilisateur (Cette liste a été élaborée en partenariat avec DataGrandEst pour la composante géo).

A noter les éléments suivants :

- Dictionnaire d'attribut : plutôt externe dans 19115. Géo2France a préféré intégrer l'aperçu du jeu de données dans l'interface du catalogue. Les plateformes trouvent qu'un dictionnaire accessible à part est suffisant.
- Propriétaire de la donnée : Il est nécessaire de mettre en avant davantage les propriétaires de la donnée. Il est intéressant d'avoir le logo du partenaire par exemple.
- Granularité des données : Cet élément est également très mal défini dans les métadonnées. Une liste des granularités possibles devrait être élaborée.

### Retex idéobfc sur les différents moissonnages entrepris (Jerome Boutet)

La plateforme IdéoBFC offre un portail avec champ de recherche unifié et une approche par thématique.

Techniquement l'intégration se fait par plusieurs canaux : un moissonnage « classique » en CSW et plus orienté opendata en CKAN (par data.gouv mais aussi certains acteurs locaux) et Poc RSS, OAI PMH (plus documentaire)

Le moissonnage qui est effectué sur la plateforme est à la fois ascendant et descendant (notamment pour moissonner les données partenaires sur datagouv en DCAT). Ce moissonnage descendant de data.gouv est plutôt efficace et automatisé, mais la remontée se fait organisation par organisation ce qui rendrait le travail difficile dans de

plus grosses structures. De manière globale, il est constaté que certaines infos ne sont pas remontées (granularité, échelle), et qu'il y a des problèmes d'hybridation.

Conclusion : Les présentations des deux plateformes ont montré des problèmes récurrents d'interopérabilité lors des moissonnages. Le GT métadonnées devra travailler à réduire ces problèmes et améliorer l'interopérabilité, notamment entre métadonnées INSPIRE et Opendata.

*Présentation autour des évolutions du Géocatalogue (Gregory Delobelle)*

Le BRGM a développé un connecteur DCAT pour interconnecter le Géocatalogue avec data.gouv. Chaque producteur aura donc à disposition un flux de métadonnées DCAT et pourra configurer son moissonnage sur data.gouv.

Un validateur qui est l'équivalent du validateur européen a également été déployé pour les métadonnées INSPIRE. Les indicateurs rapportés présentés en séance sont encore assez bas et le validateur devra aider dans leur amélioration.

A l'avenir le géocatalogue va se recentrer sur le périmètre INSPIRE, tout en permettant une interconnexion avec data.gouv.

Le registre sera également un des axes de travail. Son rôle sera discuté en commission Règles et Qualité.

La réduction de périmètre des données INSPIRE amène à la question de savoir quelle est la stratégie à adopter au niveau régional. Notamment se pose la question du format de métadonnées à utiliser. Les plateformes régionales ont globalement des infrastructures basées sur des métadonnées ISO 19115. La migration vers un moissonnage sur data.gouv devra se faire de manière progressive, avec l'accompagnement du GT Métadonnées.

Clarification sur la réduction du périmètre : Il a bien été décidé par le Ministre la réduction du périmètre des données INSPIRE aux 200+ jeux de données nationales prioritaires. Cette évolution avait été discutée en commission RMO début 2021. La liste est disponible sur le site du CNIG.

Les données locales sont néanmoins toujours soumises aux obligations OpenData et pour l'instant les obligations INSPIRE permettent d'y répondre. Le GT métadonnées pourra accompagner cette transition vers des métadonnées Open data différentes.

Le moissonnage qui se fera entre les niveaux locaux et national pourra prendre différentes formes mais un profil unique de métadonnées, élaboré par le GT métadonnées, permettra de rationaliser ces moissonnages.

Une proposition de mandat sera faite à la Commission Règle et Qualité.

La prochaine réunion sera faite sur le mapping DCAT/INSPIRE. Des travaux se feront sur Github avant la réunion (<https://github.com/cnigfr/metadonnee/tree/main/MappingINSPIRE-DCAT>).