

Compte-rendu de réunion du 2 octobre 2020

Participants :

Nom Prénom	Organisme	Présent	Excusé
Jean-Marie ARSAC	AZIMUT		X
Jean-Marie BOURGOGNE	OpenDataFrance / Validata		X
Jérôme BOUTET	Idéo Ternum Bourgogne-Franche-Comté	X	
Romain BUCHAUT	CRIGE PACA		X
Olivier DISSARD	CGDD/DRI/MIG		X
Arnauld GALLAIS	Cerema Ouest	X	
Stéphane GARCIA	IGN / Cellule normalisation		X
Clément JAMET	Métropole de Lyon		X
Marie LAMBOIS	IGN / Cellule normalisation	X	
Sébastien LAUNAY	IGN		
Mathieu Le Moal	Axes Conseil		X
Hélène LEUFROY	Géo17		X
Stéphane LEVEQUE	Cerema DTec TV	X	
Jocelyne MARC	IGN / Mission Qualité	X	
Suzanne NICEY	Idéo Ternum Bourgogne-Franche-Comté		X
Nicolas PY	IGN Centre Est	X	
Clémence RABEVOLO	IFREMER	X	
Mathieu RAJERISON	Cerema Méditerranée	X	
Gessica REYNAUD	Géomap - Imagis		X
Stéphane ROLLE	CRIGE PACA	X	
Pierre VERGEZ	IGN / Mission CNIG	X	

Ordre du jour :

- Validation du [précédent compte-rendu](#). Points d'info et d'actu.
- Une ontologie pour les métadonnées géospatiales centrées utilisateur (N. Py)
- Qualification de la précision de la BAN sur la région BFC par analyse géostatistique (J. Boutet)
- Outil de restitution de la qualification des données (M. Rajerison)
- Point divers

Prochaine réunion : **jeudi 4 février 2021**

1. Validation du précédent CR - Points d'info et d'actu

Le [compte-rendu](#) du précédent [GT CNIG QuaDoGéo](#) est relu et validé.

Le chantier « synopsis pratique pour la qualification des données » n'a pas recueilli de commentaires particuliers, ce chantier est achevé.

L'action d'abonder la nomenclature des usages génériques de l'information géographique n'a pas progressé. M. Lambois propose d'y réfléchir.

La Commission RMO du CNIG se réunira le 10 novembre. A. Gallais y dressera le bilan du GT QuaDoGéo.

N. Py fera également une présentation (en anglais) des travaux à Eurogeographics le 19 novembre.

A. Gallais annonce une possible reprise des webinaires « Data dans les territoires » en 2021.

Le groupe de travail identifie une piste de travail dans le fait d'évaluer l'expression des spécifications de qualité des données à produire dans les différents géostandards CNIG, COVADIS, ou autres..

Décision / Actions :

- *Présentations du GT CNIG QuaDoGéo en Commission RMO (A. Gallais) et à Eurogeographics (N.Py)*
- *Abonder la nomenclature des usages génériques de l'information géographique (M. Lambois)*

2. Ontologie pour les métadonnées géospatiales centrées utilisateur

Par N.Py et M. Lambois. La présentation est [disponible ici](#).

Un article de ISPRS International Journal of Geo-Information présente une action de recherche ayant permis d'établir une [ontologie](#) pour les métadonnées géospatiales destinées aux utilisateurs d'informations géographiques, afin de réduire leurs incertitudes sur la qualité tant interne que externe des données.

L'article fait le constat qu' « *alors qu'existent des méthodes pour l'évaluation de la qualité interne des données géographiques, l'évaluation de la qualité externe reste une question ouverte. Les utilisateurs, experts ou non, de données géographiques sont présumés en capacité de savoir de quel type de données ils ont besoin et où les trouver. Même avec cette information disponible, les utilisateurs ne sont pas accompagnés dans la décision de l'adéquation de cette donnée à leur besoin. Or cette décision est souvent basée sur des métadonnées complexes dans les trop peu nombreux cas où elles sont disponibles. Tandis que les efforts de standardisation ont permis l'amélioration significative de l'interopérabilité, les informations de qualité de la donnée ne sont pas transmises de manière standardisée à l'utilisateur* ».

M. Lambois présente et compare l'existant ([diapo 6](#)) en termes de normes, d'outils et de moyens de remontées d'avis d'utilisateurs dans les domaines respectifs de l'information géographique et de la donnée ouverte. [Geolabel](#) est cité en exemple pour auto-déclarer la qualité d'un lot de données.

N. Py explicite la méthode de l'étude (basée sur un questionnaire destiné à un relativement faible effectif d'utilisateurs australiens), l'analyse des résultats qui a suivi, et ses résultats traduits dans l'ontologie GUCM (Geospatial User-Centric Metadata). Celle-ci permet de décrire les données à différents niveaux de granularité : jeu de donnée, objet, attribut, en

compatibilité avec les normes ISO : 19109:2015, 19115-1:2014 et ISO 19157:2013

L'enquête révèle que, outre les critères de qualité interne, l'utilisateur final analyse également la qualité via des critères tels que : le coût ; le sentiment que l'utilisateur perçoit que la donnée est juste (sentiment lié à la confiance accordée au producteur) ; la connaissance générale de la donnée ; la pertinence de la donnée ; l'avis des pairs ; etc.

Outre la nécessité de bien décrire les jeux de données, l'ontologie GUCM préconise trois composantes : l'interopérabilité des métadonnées entre portails (à composante spatiale ou pas) ; le développement de passerelles entre les standards de métadonnées ; et la capacité à recueillir les retours utilisateurs, par exemple en utilisant le « Dataset Usage Vocabulary » (DUV) recommandé par le W3C pour publier et utiliser des données sur le web.

La suite des travaux de recherche prévoit l'implémentation de GUCM, avant de proposer un modèle finalisé et testé à l'Open Geospatial Consortium (OGC).

En conclusion de sa présentation, N. Py suggère trois idées :

- la publication, par les producteurs de données géographiques, des synthèses de recette des données par les utilisateurs (par exemple les audits externes) ;
- la création de forums utilisateurs adossés aux espaces professionnels et/ou institutionnels ;
- le maintien de registres d'utilisation des données par tel utilisateur, pour tel usage, dans tel contexte. *[où l'on retrouve le besoin d'une nomenclature des usages de l'information géographique]*

3. Qualification de la précision de la BAN sur la région BFC par analyse géostatistique

Par J. Boutet. La présentation est [disponible ici](#).

Le sujet Adresse est un sujet important pour les plateformes territoriales de données. En région Bourgogne Franche-Comté cela s'est concrétisé par une demande d'accompagnement auprès des territoires pour qualifier et améliorer la donnée adresse considérée comme d'enjeux majeur pour les politiques publiques.

IDÉO BFC a mené une brève analyse afin d'identifier les territoires où le positionnement de l'adresse dans la base adresse nationale (BAN) serait de moins bonne qualité, notamment au regard des problématiques de déploiement de la fibre avec l'enjeu de raccordement au domicile, c'est-à-dire à la « plaque adresse ».

L'étude s'est concrétisée par un traitement semi-automatique pour générer un indicateur communal défini comme le ratio entre les adresses positionnées à la « plaque adresse » et celles positionnées de façon beaucoup moins précise « à la commune ».

L'indicateur est représenté sur une cartographie interactive permettant d'identifier les territoires à sensibiliser en priorité pour une amélioration du positionnement des points adresses à travers les bases adresses locales (BAL).

Le script d'analyse développé par IDÉO BFC est basé sur des logiciels libres (R, Rstudio). Il opère sur les données ouvertes de la BAN de huit départements via son API de gestion. Il s'exécute semi-automatiquement et est disponible à la libre réutilisation sur le [github dédié](#).

La [cartographie dynamique](#) présente les différents type de localisations et permet d'identifier facilement les zones dont l'adresse devrait être améliorée, et de les croiser avec les densités communales et les conventions de déploiement de la fibre (zonages AMEL, AMII, RIP)

A. Gallais demande si cette information de qualification de l'adresse n'est pas connue des co-producteurs de la BAN (IGN, La Poste, DGFIP...) à l'issue des étapes de production, et

ne pourrait pas être transmise directement aux utilisateurs via les métadonnées.
S. Launay confirme que l'IGN procède effectivement à des calculs géostatistiques analogues.

4. Outil de restitution de la qualification des données géographiques

Par M. Rajerison. La présentation est [disponible ici](#), et le [prototype ici](#).

A l'instar de sites web du style « Les numériques » le groupe de travail s'est orienté vers une [maquette graphique](#) comprenant :

- deux notes globales : la note du testeur (vert), et celle des utilisateurs (orange) ;
- une rubrique « déposer un avis » ;
- des sous-notes suivant des thèmes (facilité d'obtention, facilité d'usage, fraîcheur, etc.) ;
- une rubrique « points forts / points faibles », potentiellement transposable en : « usages recommandés / usages non recommandés ».

M. Rajerison a entamé le développement d'un outil de génération du résultat graphique sur ces bases, en s'inspirant de l'existant à Lyon Métropole.

L'outil développé en python importe un code XML comprenant : les mesures correspondant au calcul des indicateurs de qualité interne ; des notes de synthèse de la qualité calculées depuis les métriques précédentes ; et des notes de qualité selon les usages.

L'ensemble de ces paramètres sont graphiquement générés et restitués sous forme de camemberts, feux (vert ou rouge, ou tricolore..), barre de progression, nombre d'étoiles, etc. L'outil génère une page html (report.html) intégrant les différents éléments graphiques :



Le code est libre et [disponible ici](#).

M. Rajerison soulève la question de la difficulté de définir des notes de qualité à partir des valeurs des critères (la fonction n'est pas linéaire...). L'outil révèle également le besoin d'un référentiel des usages de l'information géographique.

Il envisage plusieurs améliorations à cette preuve de concept : meilleurs rendus, superpositions de diagrammes polaires, etc.

M. Rajerison considère important qu'un portail OpenData intègre l'outil de restitution graphique de la qualité. IDÉO BFC et Crige Paca se déclarent intéressés pour tester la solution.

P. Vergez suggère un inventaire des portails géographiques qui présentent des informations relatives à la qualité des données diffusées. N. Py se propose pour cette action.

M. Lambois et N.Py recommandent de réutiliser les flux XML spécifiés : celui de la norme ISO 19157 et du Geospatial User Feedback (GUF) de l'OGC. A ce sujet, M. Lambois informe qu'une révision de ISO 19157 serait en cours (enquête en cours) prévoyant l'intégration du GUF.

Il convient également vérifier la correspondance des libellés de critères qualité entre le [registre français de mesures de qualité des données géographiques](#), et l'implémentation XML de ISO 19157.

Décision / Actions :

- Poursuivre le développement de l'outil de génération du résultat graphique conforme à la maquette et adapter le script au flux XML de la norme ISO 19157. (M. Rajerison)
- Etudier le GUF de Géonetwork, et tester la restitution sur une plateforme Géonetwork (M. Lambois)
- Faire un inventaire des IDG présentant des informations relatives à la qualité des données (N. Py)

Avancement global :

A ce stade, la production effective ou en cours du GT CNIG QuaDoGéo comprend :

- Les [fiches méthodologiques](#) du Cerema ;
- Le [registre français de mesures de qualité des données géographiques](#) ;
- La [carte mentale](#) au sujet de la qualification des données suivant la norme ISO 19157 ;
- Deux pistes exploratoires de qualification de données (synopsis, logigramme) sur des thèmes particuliers ;
- La [méthodologie](#) de qualification de données (*à consolider...*), comprenant :
 - une nomenclature des usages génériques de l'information géographique ;
 - une maquette de restitution de la qualité des données ;
- Une [présentation du GUF](#), outil de retour utilisateur quant à la qualité des données.
- (*en cours*) Le prototype d'un [outil de restitution graphique de la qualification](#) des données géographiques