

## Compte-rendu de réunion du 4 juin 2019

### Participants :

Nom Prénom	Organisme	Présent	Excusé
Jean-Marie ARSAC	AZIMUT		x
Jérôme BARET	Dept 45 / responsable SIG		x
Jérôme BOUTET	Idéo BFC - GéoBourgogne		x
Romain BUCHAUT	CRIGE PACA	x	
Hélène DURAND	Alisé Géomatique		x
Olivier DISSARD	CGDD/DRI/MIG	x	
Loïc HAY	OpenDataFrance / Validata		x
Clément JAMET	Métropole de Lyon	x	
Arnaud GALLAIS	Cerema Ouest	x	
Mathieu Le Moal	Axes Conseil	x	
Marie LAMBOIS	IGN / Cellule normalisation	x	
Stéphane LEVEQUE	Cerema DTec TV	x	
Jocelyne MARC	IGN / Mission Qualité	x	
Nicolas PY	IGN Centre Est		x
Mathieu RAJERISON	Cerema Méditerranée	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.
- Formation à distance du Cerema sur la qualification des données géographiques
- Témoignage sur les activités de Axes Conseil
- Production de synopsis pratiques pour la qualification de données
- Analyse sur les "Métadonnées sur la qualité d'un lot de données"
- Restitution des éléments de qualification d'un lot de données
- Point divers

Prochaine réunion : **18 septembre 2019 à 10h** (Saint-Mandé, IGN salle A571)

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

- Le [compte-rendu](#) du GT QuaDoGéo du 8 mars est validé.
- O. Dissard remarque que la dernière séance faisait allusion aux données en général, pas uniquement les données géographiques. Il rappelle que le rapport parlementaire sur les données géographiques souveraines marque une séparation entre les données géographiques de référence (socles) et les données métiers. Par exemple : les données géographiques métiers de l'AFB (eau, biodiversité,...), sont indépendantes des données produites par l'IGN, même si elles peuvent s'appuyer sur ce socle. Ainsi, O. Dissard recommande de garder cette distinction en tête dans les travaux du GT QuaDoGéo, afin de bien différencier entre la qualification des données métiers et celles du socle sur lequel elles s'appuient.
- Cette session accueille : J. Marc (déléguée qualité à l'IGN), M. Lambois (pilote du GT CNIG Métadonnées) et M. Le Moal du cabinet Axes Conseil (*cf présentation infra*).
- N. Py a trouvé sur [Github une consolidation de la norme ISO 19157 au format XML](#).
- Les propositions visuelles avec fiche de restitution conviviale des éléments de qualification des données ne figurent pas explicitement dans le mandat du GT QuaDoGéo mais seront néanmoins traitées comme une troisième action, complémentaire à la « Production de synopsis pratiques pour la qualification de données » et à l'analyse des « Métadonnées sur la qualité d'un lot de données ».
- En relation avec cette dernière action, A. Gallais mentionne l' « [Appel à commentaires sur le guide CNIG des métadonnées v2.0](#) » (en cours, jusqu'au 21 juin).

## 2. Formation à distance sur la qualification des données géographiques

S. Levêque en rappelle le contexte et le contenu. Le pilote est le Cerema. Cette formation est en période de test. Le déploiement est visé sur deux plateformes : [celle du Cerema](#) et celle des CVRH.

S. Rolle l'a partiellement testée pour le CRIGE PACA et considère qu'elle vulgarise bien la norme ISO 19157 mais s'adresse à un public averti. Elle s'appuie sur les fiches méthodologiques sur la qualification des données géographiques, et y ajoutant deux cas pratiques. C. Jamet suggère de se rapprocher des CNFPT pour produire un module de formation simplifiée, de type sensibilisation. S. Levêque comptait sur les CVRH pour la phase de test en visant dans un premier temps DREAL et DDT, mais cette phase n'a pas été initiée faute de répondant pour le moment.

M. Rajerison suggère le développement d'une deuxième offre de formation axée sur les possibilités offertes par l'outil QGIS en rapport avec la qualification et la correction des données géographiques (contrôles, corrections, formulaires....)

### Décision / Actions :

- ré-impulser une action de déploiement de la FOAD.
- résoudre la question de l'ouverture des supports de formation (licence et diffusion)
- étudier la possibilité d'un module de sensibilisation avec le Centre National de la Fonction Publique Territoriale.

### 3. Témoignage sur les activités de Axes Conseil

M. Le Moal présente la vision de Axes Conseil concernant les « Audits techniques de la qualité des données géographique » (cf. [cette présentation](#))

Axes Conseil intervient en assistance à maîtrise d'ouvrage SIG pour les collectivités, et développe des process qualité autour des données, sans aller dans la totalité de ce qui est demandé dans la norme ISO 19157, l'objectif étant de bien répondre à la demande plutôt que de s'engager en totalité sur la qualité au sens de la norme.

Des audits de données s'avèrent régulièrement nécessaires car nombre de données ne sont actuellement pas diffusées / diffusables (notamment en open-source) faute d'avoir été préalablement qualifiées.

Axes Conseil privilégie trois types d'analyses : la conformité OGC de la géométrie, les attributs, et la topologie de réseau, considérant que la plupart des besoins métier sont centrés dans ces domaines plutôt que, par exemple, sur la précision de position.

Les traitements (sous PostgreSQL / PostGIS opérant sur des données entrantes aux formats shape, Géodatabase, et dump PostGIS) couvrent systématiquement l'ensemble des lots de données, donc sans échantillonnage.

Axes Conseil a développé des outils permettant d'exécuter un ensemble de contrôles qualité (cf [présentation](#)). L'outil est paramétrable et s'adapte à chaque prestation par la seule définition de schémas de données, tables, attributs, etc. et il permet de traiter des données en masse comportant plus d'un millier de tables.

Le rapportage de qualité des données est composé de données issues du résultat dans une base PostgreSQL et de rapports sous forme de tableurs. Il donne une lecture neutre, sans appréciation « métier » ou thématique, également une lecture « brute » et objective sans notion de plage de tolérance. Aucune correction n'est effectuée car la décision finale revient au producteur des données qui procède à sa propre analyse avec sa vision métier. Le rapport focalise cependant son attention sur les points critiques, les tables concernées et les pistes d'améliorations emblématiques.

### 4. Production du synopsis pratique pour la qualification de données

*En rapport avec le souhait d'engager des travaux sur la construction d'un synopsis permettant de déterminer en fonction de la nature d'une donnée, quels critères / quelles mesures utiliser, un [tableau récapitulatif des critères, sous critères et mesures](#) prévus par la norme ISO 19157 avait été communiqué, et S. Levêque avait proposé une [grille d'analyse sous forme d'un tableur distant les données du SCDL \(ou autres données\)](#). L'idée consiste à remplir ce tableau avec les propositions de mesures qualité pertinentes pour chaque critère en fonction du jeu et type de données, et de les prioriser par ordre logique et d'importance dans le cadre de l'évaluation du jeu de données.*

*M. Léobet avait suggéré de se focaliser sur un jeu de données, et de produire la méthode dans le cadre d'un développement Agile : apprendre sur le sujet, déterminer si on sait le traiter ou pas, et passer à la suite... Il avait été proposé de travailler sur la [base adresse locale](#), ce qui permet de travailler en lien avec Validata, d'aborder de fait une problématique soumise à la dissémination des producteurs, et pour un enjeu qui semble bien situé dans le champ du CNIG.*

*Outre la réflexion globale sur la méthode, des expérimentations pratiques ont été*

prévues :

- Base Adresse locale : C. Jamet, L. Haÿ, N. Py, J. Boutet, J. Baret
- Equipements collectifs publics : S. Rolle et R. Buchaut, M. Rajerison
- Aménagements cyclables : S. Levêque
- Admin express : C. Jamet

### **Base Adresse Locale (BAL)**

Une séance de travail a rassemblé C. Jamet, N. Py et S. Lévêque pour traduire le synopsis de qualification des données géographiques en prenant le cas de la base adresse locale, à comprendre ici en tant que « production de la base adresse locale ».

Le brainstorming (cf. [paper-board](#), en notant que les points en vert sont ceux pour lesquels une action est nécessaire) a permis :

- d'entrevoir les questions à se poser pour mettre en place un processus qualité ;
- de dessiner les étapes et difficultés pour les utilisateurs ;
- d'écrire le cheminement permettant le processus de qualification ;
- de recenser les pistes d'actions pour le groupe.

La démarche a également permis de constater les potentielles divergences entre les points de vue du producteur dans son approche « spécifications de saisie », et celui de l'utilisateur dans une approche « exploitation sur le terrain de données open-source ».

Pour illustrer le principe, un [arbre de décision](#) (format xmind) a été développé concernant le critère d'exhaustivité, alliant les points de vue du producteur et de l'utilisateur : « - Je suis producteur ou utilisateur d'une base adresse, quelles sont les étapes pour la qualifier ? ».

La réflexion a fait ressortir la nécessité de la définition préalable rigoureuse des concepts que l'on souhaite contrôler. En l'occurrence : qu'est-ce qu'une adresse ?

Il s'agit également de bien définir ce que l'on veut partager, c'est une phase préliminaire au remplissage du tableau.

Le tableau synopsis sera par la suite complété sur la base de cette première concertation.

### **Equipements collectifs (Equipco)**

Le CRIGE PACA et le Cerema Méditerranée ont analysé la donnée « [Equipements Collectifs](#) » du [Socle Commun de Données Locales](#) (SDCL) constitué par OpenDataFrance. L'étude comporte un volet Qualification et un volet Restitution (cf [présentation](#) p1 à 9).

Le volet qualification a consisté à découvrir la donnée « Equipco ».

Une base de données locale d'équipements collectifs (comprenant 14 tables relationnelles) a été comparée au modèle de données plus généraliste du SDCL comprenant une seule table.

Après avoir déterminé les attributs annexes et indispensables, les critères et mesures qualité ont été classés par importance, en leur attribuant une note de 1 à 5 (au dessus de 3 étant considéré comme important, 5 étant le plus important).

L'analyse a confirmé qu'un critère peut s'avérer important pour un usage et pas pour un autre. A titre d'exemple les horaires d'accès sont essentiels pour un accès au public, mais pas pour une étude sur la répartition spatiale des équipements collectifs.

Ainsi, le résultat se présentera sous la forme d'une matrice présentant les critères et mesure qualité en colonnes, suivant différents usages placés en lignes.

O. Dissard propose un mécanisme s'appuyant sur une notation globale, assorties de notations secondaires correspondant aux usages présentant des exigences particulières.

Le sous-groupe a identifié quatre dimensions à intégrer pour la qualité, interconnectées entre elles : les critères et les usages (cf ci-dessus), mais aussi les attributs et les données.

La question des usages étant centrale, il semble utile de constituer un thesaurus des différents usages possibles attendus d'une donnée, et de l'intégrer dans la méthodologie de qualification.

A. Gallais ajoute que outre les dimensions « mesures qualité » et « usage » la différenciation entre l'exploitation des données par un humain et/ou par une machine est également un élément de différenciation important dans la démarche de qualification.

Décision / Actions :

- *Poursuivre les travaux, et y associer OpenDataFrance*
- *S'assurer que tous les [critères / sous-critères / mesures retenus dans les fiches méthodologiques](#) sont bien référencés dans le [tableau de qualification](#). (ex : coeff. Kappa)*
- *BAL : compléter l'arbre de décision, réaliser un test de qualification de données*
- *Equipco : Réaliser un test sur des données de Datasud pour les qualifier*
- *entamer l'action sur les Aménagements cyclables (S. Levêque) et Admin express (C. Jamet*
- *Démarrer la rédaction de la note méthodologique pratique et facile à s'approprier par l'utilisateur*

## 5. Analyse sur les "Métadonnées sur la qualité d'un lot de données"

(Voir [cette présentation](#)).

*Le sujet consiste à guider et faciliter la saisie ainsi que la transmission d'informations sur la qualité de la donnée au sein des métadonnées. Il s'agit donc d'améliorer la saisie des critères qualité dans les métadonnées ISO / Inspire en préconisant de bonnes pratiques à l'usage du producteur et de l'utilisateur. Les questions sous-jacentes seraient donc :*

- *Comment saisir des critères qualité dans les métadonnées ISO ? (fournir des consignes de saisie, par exemple pour assurer que le champ « généalogie » fournisse effectivement des informations pertinentes qualifiant la donnée) ;*
- *Comment décrire la qualité dans les métadonnées, de façon à être facilement compréhensible par l'utilisateur ;*
- *Quelle peut être la part d'automatisation de la production de ces métadonnées qualité ?*

A. Gallais s'est rapproché de l'équipe de la cellule normalisation de l'IGN (M. Lambois, S. Garcia) pour décrypter le contenu de la documentation ISO, identifier les champs concernés dans ISO19115, inventorier les éventuelles documentations existantes, et expliciter les métadonnées de qualité de façon à les rendre compréhensibles aux utilisateurs.

Des critères de qualification importants tels que l'exhaustivité ne sont pas référencés dans ISO 19115, mais les spécifications Inspire admettent les extensions de modèle compatibles ISO19115, et M. Lambois souligne d'autre part que les normes ISO 19157 et ISO 19115 doivent être considérées comme deux normes constituant un ensemble cohérent pour la transmission de métadonnées sur la qualité.

Par exemple le critère particulièrement important relatif à la « fraîcheur » des données : n'est pas porté par la norme ISO 19157 car présent dans ISO 19115. Autrement dit la norme ISO 19157 n'est pas considérée comme indépendante, mais comme devant être

utilisée avec les métadonnées ISO 19115.

La proposition du GT QuaDoGéo consiste en la création d'un registre français pour la transmission de métadonnées de qualité des données géographiques.

Ce registre sera constitué à partir :

- du sous-ensemble de la norme ISO19157 constitué des mesures retenues dans les fiches méthodologiques
- d'une extension propre aux mesures hors ISO 19157 mais préconisées dans les fiches méthodologiques (exemples : Coefficient kappa pour la justesse de classement, Classe de précision de l'arrêté de 2003 pour la précision de position)
- des trois éléments de qualité portés par 19115 : généalogie, résolution spatiale, cohérence topologique (ce dernier critère étant également présent dans ISO 19157)

Il comprendra au minimum un nom, un alias, un identifiant pour chaque élément de qualité (M. Lambois fournira un exemple)

Ce registre pourrait être porté par le site du BRGM, ou un Github, mais peut préalablement prendre la forme d'un simple document pdf.

L'outil open source [Re3gistry](#) peut être utilisé pour le constituer.

Ce registre venant comme une « extension pour les éléments de qualité » aux [Guide CNIG des métadonnées v2.0](#), les deux documents devront se référencer réciproquement.

En cours de réunion, le groupe de travail a plusieurs fois interrogé la question de l'interface homme machine pour la saisie des métadonnées en général, et des métadonnées de qualité en particulier.

#### Décision / Actions :

- M. Lambois transmet un exemple de registre d'éléments de qualité (exemple basé sur la norme ISO 19157)
- A. Gallais entame avec l'aide de la cellule normalisation de l'IGN la création du registre français des éléments de qualité dans les métadonnées.
- demander lors de l'[appel à commentaires](#) aux Guide CNIG des métadonnées v2.0 la possibilité d'une extension du guide faisant référence à un guide succinct et/ou au registre français du CNIG pour les éléments de qualité.
- étudier comment l'intégration du registre pourrait s'effectuer dans [Géonetwork](#)

## **6. Restitution des éléments de qualification d'un lot de données**

(Voir cette [présentation](#) p10 à 14)

*Le groupe de travail préconise l'élaboration d'une fiche de restitution « conviviale » à la fois compréhensible et agréable à consulter : c'est-à-dire une synthèse abondamment graphique du contenu d'un rapport qualité, et y associant la possibilité d'un accès direct en cas de besoin d'informations plus détaillées.*

*On élaborera simultanément des méthodes de restitution « conviviale » basées sur des aspects communication, vulgarisation, pédagogie en utilisant autant que possible le langage graphique à l'aide d'éléments constituant un juste compromis entre l'indicateur synthétique (feu tricolore, etc...) et la référence technique à la norme ISO 19157.*

Il est important que la restitution de la qualification des données facilite la compréhension de l'utilisateur en expliquant certaines mesures assez complexes (telles que celles

d'incertitude, les coefficients kappa), soit sous forme textuelle, soit imagée grâce à des illustrations et pictogrammes.

La restitution peut apparaître sous plusieurs formes :

- synthétique, sous forme de diagramme polaire faisant apparaître les notes affectées à chaque critère
- ou plus détaillée faisant apparaître les notes pour chaque mesure.

Il est possible et intéressant d'affecter une pondération aux différentes mesures d'un critère.

S'il semble relativement aisé d'évaluer la qualité d'un jeu de données [en fonction des usages], il apparaît en revanche assez difficile de noter chaque sous-critère, critère, et enfin l'ensemble du jeu de données. Cet aspect méritera d'être étudié lors d'une prochaine session.

M. Rajerison formule les premières propositions visuelles basées sur des « pictos qualité » exprimant visuellement la nature du concept, et des « vu-mètres » pour la notation (de 1 à 5) afin de présenter une fiche de restitution conviviale.

Il apparaît fondamental que plusieurs niveaux de lecture soient proposés à l'utilisateur. A cet effet des « résultats dépliés » permettent des restitutions en cascade du moins au plus détaillé.

Enfin, une remontée des impressions utilisateur (mécanisme bottom-up) semble intéressante à implémenter dans les portails de diffusion de données, afin de prendre connaissance des différents usages de la donnée, et des difficultés rencontrées dans leur utilisation.

A ce sujet, le standard [Geospatial User Feedback](#) (GUF), utilisé dans certains projets européens, méritera d'être étudié au sein du groupe.

#### Décision / Actions

- *Poursuite de l'analyse et de propositions visuelles pour une fiche de restitution conviviale des éléments de qualification des données. ( S.Rolle, R. Buchaut, M. Rajerison)*
- *Analyse des apports de Geospatial User Feedback*