

Compte-rendu de réunion du 8 mars 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	
Raphael DELHOME	Oslandia		x
Olivier DISSARD	CGDD/DRI/MIG		x
Tony EMERY	CCPRO		x
Loïc HAÏ	OpenDataFrance / Validata	x	
Clément JAMET	Métropole de Lyon	x	
Didier JOSSELIN	Université Avignon		x
Arnauld GALLAIS	Cerema Ouest	x	
Marc LEOBET	CGDD/DRI/MIG	x	
Stéphane LEVEQUE	Cerema Centre Est	x	
Stéphane MEVEL-VIANNAY	GéoBretagne		x
Nicolas PY	IGN Centre Est	x	
Mathieu RAJERISON	Cerema Méditerranée	x	
Elodie ROBINET	SOLURIS		x
Stéphane ROLLE	CRIGE PACA	x	
Pierre VERGEZ	IGN / Mission CNIG	x	

Ordre du jour :

- Revue du [mandat du GT CNIG QuaDoGéo](#), objectifs, fonctionnement du groupe
- Points d'info / actu / communication (QuaDoGéo les épisodes précédents)
- Présentations et retours d'expériences locales
- Production collective d'un synopsis pratique pour la qualification de données
- Organisation pour la production d'une fiche "Métadonnées sur la qualité d'un lot de données"

Prochaine réunion : **4 juin 2019 à 10h** (IGN saint-Mandé, salle A670)

1. Revue du mandat du GT CNIG QuaDoGéo, objectifs, fonctionnement du groupe

A. Gallais parcourt en séance le [mandat du GT CNIG QuaDoGéo](#), validé par la Commission Règles de mises en œuvre (RMO) du CNIG le 7 décembre 2018.

Le sujet de la qualité des données géographiques est un sujet ancien et connu, mais encore relativement mal cerné. Il a paru nécessaire de constituer un groupe de travail pour développer l'activité de qualification des données géographiques afin de produire si possible des solutions concrètes.

Le Cerema a récemment travaillé sur le décryptage de la norme ISO 19157. Celle-ci s'intéresse principalement à la qualité interne, en adoptant le point de vue du producteur de données géographiques. Le groupe de travail du CNIG s'intéressera également à la qualité externe et au point de vue de l'utilisateur.

Il est rappelé l'intérêt pour le groupe de travail d'associer différentes parties prenantes : collectivités et leurs représentants, producteurs de données, éditeurs de logiciels, entreprises privées, labos de recherche, etc.

Le groupe de travail se donne pour mission d'aider à la réutilisation des données ouvertes grâce à l'information sur la qualité.

Contrairement à la norme, le « retour utilisateur » sera pris en compte. Sa participation à l'évaluation et à la qualification des données était une idée prégnante du [séminaire sur la qualification des données géographiques](#) qui s'est tenu à Lyon en février 2018.

M. Léobet, président de la Commission RMO du CNIG, a souhaité que le GT QuaDoGéo prévoit un point d'étape à échéance d'un an afin d'évaluer sa production. Cette règle s'applique classiquement aux GT du CNIG afin de s'assurer de leur utilité et de hiérarchiser leurs objectifs. L'important dans un GT du CNIG est de produire des livrables à horizon d'un an dans le cadre de la méthodologie commune du CNIG basée sur des espaces de concertation ouverts débouchant sur des réalisations (géostandards, etc.) ensuite soumises à consultation publique sous forme d'un « appel à commentaires ».

Ainsi, deux actions concrètes réalisables à horizon d'un an sont prévues dans le mandat :

- la confection d'un « synopsis pratique pour la qualification » : il s'agirait, face à un type de données, de déterminer une méthode et un ordre des critères à mesurer ;
- la réalisation d'une première fiche qui validera la méthode. Le groupe de travail déterminera le processus à mettre en place : type de fiche, choix éventuel d'un lot de données test, etc.

Débat :

- L. Haÿ confirme qu'il y a bien deux publics ciblés par la qualification des données : producteur et utilisateur, et que les données peuvent être géographiques ou pas. Le GT s'accorde à traiter en priorité les données géographiques mais sans s'y cantonner compte tenu de l'essor des réalisations d'Open Data France et l'écosystème de l'Open-Data.

- N. Py propose une troisième action consistant en l'établissement d'un glossaire des principaux concepts du domaine : qualité interne, qualité externe, données souveraines, etc. afin que l'ensemble de la communauté adopte un vocabulaire commun. Dans le même ordre d'idée (de vulgarisation), il suggère également la conception et réalisation d'actions pédagogiques à destination des étudiants et de centres de formations.

- P. Vergez pose la question de la nature du lien entre le GT CNIG QuaDoGéo et la Commission Géopositionnement au sujet de l'[arrêté du 16 septembre 2003 portant sur les classes de précision applicables aux catégories de travaux topographiques](#). : Le GT QuaDoGéo n'est pas mandaté pour travailler à la révision de cet arrêté. P. Vergez assurera cependant le lien entre le GT QuaDoGéo et la Commission Géopositionnement du CNIG qui est et reste l'instance experte dans le domaine.

Décision / Actions

- Le mandat est approuvé ainsi que l'objectif de réalisation à horizon un an. Le groupe se réunira au rythme d'environ quatre réunions par an. A l'instar d'autres GT du CNIG l'activité pourra être répartie en sous-groupes pour d'éventuels sujets si nécessaire.

- J. Desboeufs (Etalab, beta.gouv.fr, geo.data.gouv.fr), et M. Lambois (cellule normalisation de l'IGN) seront invités à participer au groupe de travail.

2. Points d'info / actu / communication « QuaDoGéo les épisodes précédents »

1) Historique

A. Gallais dresse [un rapide historique du sujet de la qualité des données géographiques](#), occasion de souligner les apports successifs et conjoints de l'IGN, du CERTU (désormais devenu un service du Cerema), et du CRIGE PACA en la matière.

Les dernières étapes de cet historique sont marquées par :

- La [journée technique sur la Qualité](#), organisée par le CRIGE PACA en juin 2017.
- la publication d'une [série de fiches Cerema](#), intitulée : « *Qualifier les données géographiques - Un décryptage de la norme ISO 19157* » ;
- l'organisation conjointe par le CRIGE PACA et le Cerema [d'un séminaire en février 2018](#), duquel a émergé l'idée de la création du GT CNIG QuaDoGéo ;
- la réalisation d'une enquête auprès des utilisateurs, dont [la synthèse a été publiée en juin 2018](#) ;
- l'élaboration par le Cerema (suivie du test par le CRIGE PACA) d'une formation à distance (FOAD) sur la qualification des données géographiques, basée sur les fiches méthodologiques et augmentée de deux études de cas plaçant l'apprenant en situation réelle de qualification de données géographiques ;
- deux réponses au programme d'investissements d'avenir (PIA) 2018 dans le domaine de la qualité des données : Qualidon et Qualidata, le deuxième ayant été retenu et se nomme désormais [Validata](#) (cf. §2.5) ;
- Présentation par le Cerema de QuaDoGéo au Géodatadays 2018

2) Communication

M. Rajerison dresse le panorama des médias de communication développés sur le sujet :

- [rubriques en relation avec QuadoGéo](#) sur le site du Cerema ;
- fil de discussion « [Qualité des données géographiques : projet QuadoGeo](#) » sur Géorezo ;
- un [compte](#) et un [hashtag Twitter](#) ;
- sur un plan plus institutionnel, s'y ajoute désormais [la page du GT CNIG QuaDoGéo](#).

A mentionner également les travaux de [#TeamOpenData](#), d'autant plus intéressants qu'ils

ne sont pas restreints au domaine de l'information géographique.

Le groupe de travail se propose de communiquer largement, également auprès de l'Afigéo et l'AITF.

3. Présentations et retours d'expériences locales

1) IDéOBFC

J. Boutet présente la perception de la qualité en région Bourgogne Franche-Comté. ([présentation](#)). Il signale un début de réflexion de IDéOBFC avec ses partenaires, et souligne la disparité des acteurs sur le sujet. La question de la qualité des données géographiques est bien reconnue comme un sujet important mais face à la dualité entre « diffusion de données géographiques en l'état de leur qualité » et « pas de diffusion avant (re)mise en qualité à réaliser », les actions de correction et de diffusion s'avèrent le plus souvent repoussées...

IDéOBFC insiste sur la nécessité de bien saisir les métadonnées, notamment la balise généalogie, afin de répondre en partie aux besoins de réutilisation, et fait un parallèle avec sa perception de l'Open-data où la diffusion peut éventuellement précéder l'amélioration du lot de données.

IDéOBFC se propose de se faire le relai du GT QuaDoGéo en région BFC, de diffuser les productions et de sensibiliser la communauté d'utilisateurs sur les fiches méthodologiques.

2) CRIGE PACA

Le CRIGE PACA œuvre depuis longtemps dans le domaine de la qualité des données géographiques (cf. §2 et [cette présentation](#)).

S. Rolle et R. Buchaut présentent une [vidéo de sensibilisation en 2mn vraiment très intéressante !](#) au sujet de la qualité des données géographiques, récemment réalisée par le CRIGE PACA.

Le CRIGE réalise des actions d'acculturation (journées techniques) de communication, d'assistance et enseignement auprès de ses partenaires. Il participe naturellement au bêta-test de la FOAD (formation à distance) dans le cadre de son partenariat avec le Cerema¹.

En interne, le CRIGE PACA met en place des processus automatiques souvent basés sur FME permettant notamment : l'appariement de données sur le filaire de la BD Topographique, le contrôle de conformité aux géostandards CNIG PLU, le rapport qualité sur flux de données WFS (permettant par exemple d'éditer automatiquement des comparaisons inter-départementales sur certains critères qualité), etc.

Débat :

L. Haÿ souligne la difficulté d'extraire de l'information normalisée depuis un système d'information, et s'enquiert de la possibilité de partager et diffuser les outils et ressources techniques ainsi développées.

A. Gallais constate l'avance de la région PACA en matière de qualité des données géographiques et demande quelles autres régions ont bien avancé dans le domaine. Le CRAIG (Auvergne-Rhône-Alpes) est cité pour ses développements sur la qualité interne.

¹ La FOAD sera ensuite testé au sein des centres de formation du Ministère.

3) Département du Loiret

J. Baret indique que son département travaille en réseau avec ceux de la région Centre. Dans le Loiret, le syndicat mixte numérique accompagne les collectivités locales dans leur transformation numérique. Les données (principalement les thématiques routes assainissement, eau, etc.) sont publiées sur la plateforme départementale Géoloiret.

J. Baret entrevoit la problématique de la qualité des données géographiques sous l'angle de la culture des risques naturels, historiquement importante dans la région du fait des crues de la Loire. En s'appuyant sur l'exemple d'un exercice de crise en novembre 2018, il signale les problématiques liées à la duplicité de la donnée (plusieurs sources), à son absence (les écoles), à sa qualité (absence de certaines lignes et d'arrêts de transport scolaires) et plus généralement au défaut de données estampillées « de référence ».

J. Baret s'intéresserait à la création d'« indice de confiance » ou de matrices d'« indices de confiance », notamment dans l'utilisation de données géographiques en situation de gestion du risque.

Débat :

C. Jamet privilégie la qualification et la fourniture de critères qualité objectifs, à l'idée de se baser sur un (des) indice(s) de confiance.

M. Léobet n'est pas opposé à l'utilisation d'indice de confiance, à condition qu'il soit bien décliné par types d'usages, afin d'offrir uniquement un 1^{er} niveau de réponse sur la qualité de la donnée.

L. Haÿ remarque que le responsable et/ou producteur ne fait pas nécessairement référence en vertu de son statut (cf cas des écoles ci-dessus).

4) Métropole de Lyon

La qualification de données a permis à la Métropole de Lyon de constater des défauts d'exhaustivité sur la BD Adresse produite. Il s'en est suivi une remise en qualité assortie de l'utilisation de la fiche « échantillonnage » permettant d'organiser une partie des contrôles sur le terrain.

Un travail conséquent est en cours dans le SI de la Métropole de Lyon afin de simplifier et d'homogénéiser les procédures et les outils liés à la gestion du circuit des données. Les notions de qualité et de qualification des données sont pleinement intégrées dans les processus qui s'appliquent tout au long du cycle de vie des données.

Dans un premier temps, la Métropole de Lyon va - dans le cadre de l'Open-Data - publier une fiche qualité annexe aux différents lots de données. Dans certains cas précis, la publication des algorithmes de calcul des critères qualité est envisagée.

C. Jamet souhaite engager des travaux sur la construction d'un tableau / synopsis permettant d'anticiper et diriger le contrôle des données. Cette idée émane également du séminaire de février 2018. Il s'agirait d'un modèle permettant en fonction de telle donnée ou type de donnée d'indiquer les mesures nécessaires à l'évaluation de sa qualité, et la manière d'intégrer les résultats des mesures dans les métadonnées.

5) OpenData France

L. Haÿ rappelle les enjeux d'harmonisation des données diffusées : l'assurance qualité et la facilité d'utilisation / réutilisation. En d'autres termes : « il faut en finir avec le capharnaüm de données semblables sur des modèles différents ».

A travers l'élaboration et la publication du [socle commun des données localisées](#) (SDCL) OpenData France a identifié les thématiques prioritaires et/ou produits universellement utilisés par les collectivités locales, et leur a proposé d'élaborer collectivement des

spécifications et implémentations standardisées.

Les implémentations ont été décrites dans le format « Table Schema frictionless data » de l'Open Knowledge Foundation. La documentation du format « Table schema » est auto-générée, elle est lisible par l'humain, tandis que les schémas techniques sont exploitables par les algorithmes.

Le [projet Validata](#) assure le volet vérification de la structure des jeux de données produits. Validata est une « *plateforme de validation de données ouvertes produites par des acteurs territoriaux. Elle offre à ses utilisateurs la possibilité de vérifier la conformité de ces données par rapport aux modèles définis dans le cadre du SCDL. En détectant d'éventuelles erreurs ou anomalies, en amont ou en aval de leur publication, elle aide à harmoniser les données locales ouvertes, elle engage à améliorer leur qualité et elle facilite leur exploitation par les réutilisateurs* ».

L. Haÿ constate que les formulaires de saisie alimentés par un modèle d'une part, et l'existence d'outils de validation² d'autre part facilitent largement l'adoption d'un standard.

Le SCDL couvre actuellement : [la base adresse locale](#) (BAL), [le catalogue simplifié des données opendata](#) diffusées par une collectivité, les [délibérations](#) adoptées par une collectivité locale, les [équipements](#) collectifs publics, les infrastructures de recharge de véhicules électriques ([IRVE](#)), les [marchés publics](#) attribués par une collectivité locale, les [prénoms des nouveaux nés](#), les [subventions attribuées](#), etc. : ces données ne sont donc pas nécessairement géographiques.

L. Haÿ demande de lui transmettre les validateurs à notre connaissance (*remarque : fait hors réunion par N. Py pour le validateur GPU*) de façon à les intégrer à la rubrique [Validateurs externes](#) de Validata.

Outre ces outils, L. Haÿ a transmis au GT QuaDoGéo des ressources utiles et inspirantes au sujet de la qualité des données :

- le « [petit guide de survie](#) » utile au « *balayage systématique des problématiques pratiques de la qualité afin d'apporter des pistes et méthodes d'amélioration concrète* » ;
- ainsi que le « [Sprint qualité](#) » visant « *l'amélioration significative de la qualité d'un jeu de données* ».

4. Production collective d'un synopsis pratique pour la qualification de données

En rapport avec le souhait du Grand Lyon 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 (cf §3.4), M. Rajerison communique un [tableau récapitulatif des critères, sous critères et mesures](#) prévus par la norme ISO 19157, et S. Levêque propose une [grille](#) d'analyse sous forme d'un tableur éditable basé sur des données du SCDL (ou autres données), inscrivant en ligne le produit considéré et en colonnes les "indicateurs" retenus et décrits dans les fiches méthodologiques. 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 suggère 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

² à l'instar par exemple du validateur GPU pour les documents d'urbanisme dématérialisés suivant le géostandard CNIG PLU

ou pas, et passer à la suite... Il propose 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 (davantage que celui du positionnement des défibrillateurs, par exemple).

S. Lévêque rappelle l'objectif du mandat de travailler simultanément sur une méthode générale et de l'appliquer à quelques cas particuliers, si possible encadrés par l'existence d'un géostandard (permettant une première séries de qualifications suivant le critère de cohérence logique). S. Lévêque et A. Gallais souhaitent que le GT puisse s'organiser pour avancer entre les réunions.

Décision / Actions :

- *Réflexion sur la méthode globale sur la base du tableau présenté en séance : tous*

- *Cas pratiques :*

- *Base Adresse locale : C. Jamet, L. Hajj, N. Py, J. Boutet, J. Baret*

- *Equipements collectifs publics : S. Rolle et R. Buchaut, M. Rajerison*

- *Aménagements cyclables : S. Levêque*

- *Admin express : C. Jamet*

5. Organisation pour la production d'une fiche "Métadonnées sur la qualité d'un lot de données"

a) Saisie des métadonnées de qualité

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 ;
- Comment restituer (lisiblement) les métadonnées qualité à l'intention de l'utilisateur ?

M. Léobet rappelle les guides Inspire et les normes utiles : ISO19113, 114, 115, 131 138, 157, etc. Il signale que trois éléments essentiels de description de la qualité sont déjà prévus dans [le guide CNIG de saisie des éléments de métadonnées Inspire](#) : généalogie, résolution spatiale, cohérence topologique.

D'autres critères importants tels que l'exhaustivité n'y sont pas présents, mais les spécifications Inspire admettent les extensions de modèle compatibles ISO19115.

Or ISO 19115 « *permet la définition de plusieurs ensembles d'information de qualité (...)* pouvant fournir des informations de généalogie et des résultats d'évaluation de la qualité » ([page 99 du guide CNIG](#)). Il y aurait donc possibilité d'adjoindre aux métadonnées un document spécifique à la qualification des données et/ou au résultat d'un contrôle qualité.

M. Léobet recommande de se rapprocher de l'équipe de la cellule normalisation de l'IGN (M. Lambois) 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. En tout état de cause l'expertise de la cellule normalisation de l'IGN est susceptible de fournir une précieuse assistance.

b) Restitution d'informations de qualité à partir des métadonnées de qualité

C. Jamet préconise pour la phase aval 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.

Ces deux actions (consignes de saisie, restitution des métadonnées) sont fortement liées :

- On s'intéressera aux champs exploitables de la norme ISO 19115 en vue de transmettre les critères et mesures indiqués dans les fiches méthodologiques, notamment ceux traductibles en note attribuée à tel critère qualité du lot de données.
- 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.

Décision / Actions

- *A. Gallais se rapproche de la cellule normalisation de l'IGN (M. Lambois) afin d'évaluer l'existant ISO et inspire des métadonnées sur la qualité afin de les compléter et utiliser les champs comme vecteur de transmission des éléments de qualification des données.*
- *S.Rolle, R. Buchaut, M. Rajerison formulent les premières propositions visuelles (diagrammes polaires, smiley, étoiles, feux tricolores, etc) d'une fiche de restitution conviviale des éléments de qualification des données.*