

Compte-rendu de réunion du 21 novembre 2019

Participants :

Nom Prénom	Organisme	Présent	Excusé
Jean-Marie ARSAC	AZIMUT	x	
Jérôme BARET	Dept 45 / responsable SIG	x	
Jean-Marie BOURGOGNE	OpenDataFrance / Validata	x	
Jérôme BOUTET	Idéo BFC - GéoBourgogne		x
Romain BUCHAUT	CRIGE PACA		x
Chantal COULOMB	IGN / Contrôle Qualité		x
Olivier DISSARD	CGDD/DRI/MIG	x	
Clément JAMET	Métropole de Lyon	x	
Arnauld 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	
Gessica REYNAUD	Géomap-Imagis	visio	
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.
- Qualification des données d'urbanisme publiées sur le GPU (*Géomap-Imagis*)
- OpenData France et qualité des données (*J-M Bourgogne*)
- Registre français pour les métadonnées relatives à la qualité des données géographiques (*A. Gallais*)
- Remontées utilisateurs : GUF et sphère opendata (*M. Rajerison*)
- Production de synopsis pratiques pour la qualification de données
 - Base adresse locale (*S. Levêque*)
 - Equipco (*S. Rolle*)
- Restitution de la qualification des données (*M. Rajerison*)
- Point divers
 - FOAD, session test (*S. Levêque*)

Prochaine réunion : **7 février 2020 à 10h à la Maison de la Bretagne, Paris**

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

- Le [compte-rendu](#) du [GT QuaDoGéo](#) du 18 septembre est relu et validé. La plupart des actions prévues ont progressé ou été réalisées.

Au sujet de la norme ISO 19157 sur Github, N.Py suggère la formulation : *Il existe une consolidation de la norme [ISO 19157 au format XML](#) sur le dépôt [GitHub officiel de l'ISO/TC 211](#). Les schémas liés aux standards d'informations géographiques sont d'ailleurs [accessibles ici](#). L'action : « étudier / exploiter la consolidation de la norme ISO 19157 au format XML le dépôt [GitHub officiel de l'ISO/TC 211](#) » reste d'actualité.*

N. Py remarque que l'outil Ripart était cité dans le précédent compte-rendu. Il rappelle qu'il est davantage conçu pour saisir des alertes d'évolutions sur le terrain que pour effectuer de véritables retours sur la qualité des données, et informe que les outils collaboratifs de l'IGN, dont faisait partie Ripart, ont été fondus dans [l'espace collaboratif IGN](#).

M. Le Moal signale la publication d'un [article dans Techniques Sciences Méthodes](#) (TSM)

O. Dissard signale que la MIG va être fusionnée au sein d'un Ecolab.

Décision / Actions :

- *Améliorer l'aspect collaboratif du GT CNIG QuadoGéo en proposant une page de ressources sur le site du CNIG (A. Gallais, via P. Vergez), et en étudiant l'intérêt de créer un Github QuaDoGéo (M. Rajerison, S. Rolle)*

2. Qualification des données d'urbanisme publiées sur le GPU

Par G. Reynaud ([GEOMAP-IMAGIS](#)).

« Une problématique de qualification des données géographiques est apparue lors du [GT CNIG DDU](#) du 12 septembre : le validateur GPU ne contrôlant que certaines caractéristiques des lots de données les utilisateurs du GPU s'étonnent de télécharger des données dont la qualité ne correspond pas à celle attendue malgré le fait qu'elles aient été « validées par le GPU » (le terme de « validateur » entretient une confusion avec la qualification des données). »

La présentation de G. Reynaud est [accessible ici](#). Elle développe le processus d'intégration des données d'urbanisme téléchargées du GPU, insiste sur le fait que les non-conformités font échouer ce processus, et qu'il conviendrait d'indiquer que les exigences du GPU ne sont pas aussi élevées que celles des standards sur lesquels il s'appuie.

Débat :

- A. Gallais traduit les constats de Géomap-Imagis en anomalies relatives principalement aux critères de cohérence logique et topologique.

- S. Rolle indique que la réglementation impose la publication des documents d'urbanisme dans le GPU au standard CNIG, rôle dont il s'acquitte parfaitement. Dès lors il serait erroné d'indiquer que le GPU ne serait pas conforme.

- O. Dissard recommande de tenir compte de la dimension temporelle : non seulement les standards évoluent, mais les documents d'urbanisme publiés il y a 5 ans ne respectent plus nécessairement les contrôles appliqués actuellement par le validateur GPU.

- N.Py rebondit sur la question de la temporalité en rappelant que la montée en charge du GPU porte sur 25000 documents d'urbanisme et s'étale sur plusieurs années, ce qui définit le niveau d'exigence. Ainsi, le rôle du validateur reste bien de vérifier la structuration du document plutôt que la parfaite conformité au standard CNIG. A titre d'illustration, des contrôles de topologie accrus correspondraient à un niveau d'exigence supplémentaire. Si une montée des exigences graduelle au fil du temps reste envisageable, N. Py fait remarquer que si la parfaite conformité était maintenant recherchée se poserait la question de la validation du stock de documents préalablement publiés. Finalement, ce cas d'école rencontré dans le domaine de l'urbanisme semble assez général à la création des applications et l'évolution de standards sur lesquels elles s'appuient.

Décision / Actions :

- G.Reynaud se rapprochera du projet GPU pour lui proposer sa participation à la recette prévue du validateur GPU dans la prochaine version attendue du GPU.

3. OpenData France et qualité des données

Par J-M Bourgogne, Délégué Général de OpenDataFrance

La présentation est [accessible ici](#).

A l'occasion de la date anniversaire de la loi république numérique, promulguée le 8 octobre 2016, OpenDataFrance poursuit la mise à jour de son observatoire réalisé avec le concours d'Etalab et de la Banque des Territoires.

L'analyse des indicateurs confirme deux faits marquants sur l'ouverture des données dans les territoires.

- Le premier concerne l'engagement des collectivités territoriales dans l'open data avec plus de 600 acteurs locaux ayant publié leurs données, ce qui en fait le pays le plus avancé au niveau international sur cette dynamique. Il faut cependant remarquer que si la majorité des grosses collectivités sont engagées dans l'open data (région, département, métropoles), les communes et EPCI de moins de 100000 hab. restent très largement en retrait de cette démarche (moins de 8%). Cela confirme la difficulté perçue par ces acteurs et la nécessité d'un accompagnement de proximité fort pour aider les acteurs locaux à comprendre les enjeux de transparence et d'innovation de l'open data.

C'est précisément l'objectif du programme OpenDataLocale qui s'appuie sur cinquante-cinq collectivités de niveau supérieur pour soutenir les petites et moyennes collectivités. Renouvelé pour la saison 2019-2020, ces « ambassadeurs locaux de l'open data » proposent une palette de services tels que la sensibilisation, la formation à la mise en œuvre, le coaching de proximité, l'hébergement des données sur un portail mutualisé, des outils de production de qualité, l'animation territoriale, etc.

On constate une progression constante mais encore trop faible du nombre de collectivités qui ouvrent leurs données : 5200 collectivités sont concernées, il faudrait 35 ans au rythme actuel, ce qui impose d'accélérer le mouvement...

- Le deuxième constat concerne la nature et la qualité des données. Elles sont publiées dans de nombreux endroits : les portails locaux, le référencement national data.gouv.fr, quelques portails thématiques (transport.data.gouv.fr, ..) ou d'agrégation régionale (DataSud), on dénombre aujourd'hui 150 portails... Très peu respectent les standards établis qui existent maintenant depuis près de trois ans, sous l'initiative de l'état

(schema.data.gouv.fr) et des collectivités (scdl.opendatafrance.net). On remarque que sur 24000 jeux de données, moins de 4 % respectent un format défini. Par ailleurs les métadonnées restent insuffisamment renseignées. Des informations comme la fréquence de mise à jour, la licence appliquée, la thématique utilisée, sont documentées de façon très incomplète et inégale. La production et la documentation des données publiques doivent donc être largement améliorées et cet effort ne doit pas se faire au détriment des collectivités.

Malgré la présence de 54 animateurs territoriaux des données (ATD), la qualité des données reste un frein à l'utilisation et à la crédibilité de l'opendata. OpenDataFrance a mis en place une plateforme appelée [OpenDataFactory](#) qui offre des outils méthodologiques et techniques pour faciliter la production de données de qualité : élaboration de standards simples ([espace de publication des formats stabilisé](#) et [espace de production](#)), logiciel de validation (go.validata.fr : effectue des contrôles sur les champs de données au format « Good Tables » et fournit une indication de la qualité : vert / orange / rouge), outils d'extraction des données auprès des éditeurs métiers (OpenDataButtonInitiative dont l'objectif est de générer des bases de données dans des standards de données ouvertes depuis des logiciels métier et/ou propriétaires) ou d'OpenStreetMap (Oser-OSM), application en ligne d'aide à la production de données de qualité ([ChampLibreCollecte](#)).

Devant la multiplicité des différents référentiels qui voient le jour (SCDL et <https://schema.data.gouv.fr/> vu précédemment mais il y en a d'autres) et la multiplicité des producteurs de standards (Etalab, OpenDataFrance, CNIG, ministère de la santé, Afigeo, autres acteurs privés comme METIS, etc.), il est très important de faire mûrir encore l'écosystème « data public » en organisant la gouvernance des standards, des plateformes, des outils, des méthodes. Il sera question ici de mettre en place une plateforme technique unique, des règles de gestion des ressources, l'animation des parties-prenantes. Prochaine réunion avec Etalab le 25 novembre pour un début de mise en œuvre à partir de janvier 2020.

OpenDataFrance confirme son intérêt pour toute collaboration avec le CNIG sur les sujets : standards et gouvernance.

4. Registre français pour les "Métadonnées relatives à la qualité des données géographiques"

Par A. Gallais. [Le projet de registre est accessible ici](#).

Le sujet consiste à guider et faciliter la saisie et la transmission d'informations sur la qualité de la donnée en créant des rubriques de métadonnées dédiées. Il s'agit 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 sont :

- *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 ;*

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 est 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)
- et des trois éléments de qualité portés par 19115 : généalogie, résolution spatiale, cohérence topologique (ce dernier critère étant également présent dans ISO 19157)

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.

A. Gallais présente le [projet de registre français des métadonnées de qualité des données géographiques](#). Ce registre prend préalablement la forme d'un simple document pdf avant d'être implémenté sur le « [Système de publication de registres](#) » du [Géocatalogue](#).

La nouvelle version du registre comprend l'ajout des rubriques suivantes :

- [taux d'erreur de formatage](#) §2.2.3 ;
- étendue temporelle, et dates de référence (conformes aux métadonnées INSPIRE) permettant d'intégrer le critère d'[actualité / fraîcheur](#) du lot de données ;
- exactitude de position et de radiométrie des données [matricielle](#) (extraite de la norme ISO 19157), par exemple pour les ortho-photos.

N. Py questionne la pertinence du coefficient Kappa et d'autres choix de mesures de la norme ISO 19157. Il est rappelé que ces mesures de qualité ont été sélectionnées lors de la démarche et la rédaction des [fiches méthodologiques](#).

A. Gallais a testé avec l'assistance du BRGM l'intégration d'une partie du registre (limitée aux 4 premières mesures) dans le « Système de publication de registres » du Géocatalogue : <https://data.geocatalogue.fr/ncl/mesuresQuaDoGeo>. Ces débuts sont prometteurs quoique de nombreuses améliorations soient encore nécessaires afin que toutes les caractéristiques des mesures soient disponibles.

A terme, chaque mesure sera référencée par une URL et les outils de qualification des données (tels que : Validata, validateur GPU, etc...) devront se référer au registre.

Décision / Actions :

- *relecture et commentaires en vue de la validation à la prochaine réunion (tout le GT)*
- *poursuite de l'intégration dans le « Système de publication de registres » du Géocatalogue (A. Gallais avec l'assistance du BRGM)*

5. Remontées des utilisateurs sur la qualité : GUF et sphère opendata

Par M. Rajerison. La présentation est [accessible ici](#).

La qualité interne s'appuie sur des spécifications (types de champs, contraintes, formatages) tandis que la qualité externe s'appuie sur des usages (enquête terrain, statistiques générales).

Une donnée de mauvaise qualité interne ne présente pas nécessairement une mauvaise qualité externe. A titre d'exemple, une donnée ne répondant pas à ses spécifications de

qualité sur, par exemple, la position, peut en revanche s'avérer utile pour un usage à petite échelle pour de la statistique supra-communale.

Comment évaluer ces composantes de la qualité ? S'il existe des outils permettant de contrôler la qualité interne de données (comme [Validata](#), le validateur GPU, etc.) on observe, dans le paysage des portails OpenData, des implémentations plutôt rudimentaires de la remontée d'anomalies orientée utilisateur alors que cette dernière permettrait de qualifier la donnée selon les usages, sachant que tous ne sont pas déterminables à l'avance par le producteur (principe de [sérénipidité](#)) : l'usage peut être totalement différent in fine, du fait que l'on ne puisse imaginer tous les usages à l'avance

Le [Geospatial User Feedback](#), une norme de l'OGC, définit en des termes opérationnels : le modèle conceptuel de données, le greffon applicatif, la façon d'effectuer le retour usager essentiel à la qualité externe. La question du "fitness for purpose" (aptitude à l'emploi) est placée au cœur du GUF, quoique son modèle reste complexe et mériterait d'être simplifié.

Les portails pourraient s'appuyer au moins sur une version allégée de celle-ci pour collecter les usages faits des données. Or actuellement, le retour utilisateur, faute de disposer des canaux suffisants à sa remontée, s'effectue via les réseaux sociaux, de façon parfois virulente, au détriment de l'image des institutions productrices de données. Ainsi, Twitter s'impose parfois comme le « wall of shame », faute d'existence de canaux officiels de remontée des défauts de qualité.

Les réutilisations, quant à elles, restent trop souvent associées à une haute technicité alors que leur intérêt premier est de faire part d'un usage. Or, le développement des usages, tout comme de leur connaissance (pourquoi ne pas imaginer un observatoire des usages) est le leitmotiv de l'Open Data. La qualité des données, ou dans une moindre mesure leur connaissance, autant interne qu'externe, permettrait de développer les usages car elle permettrait de s'affranchir, ou du moins alléger les étapes de qualification côté producteur, requalification côté utilisateur, voire mise en conformité dans certains cas.

En allant plus loin, la production de données certifiée selon des standards bien établis comme ceux du CNIG, OpenDataFrance ou d'Etalab, optimise le cycle de la vie de la donnée, et favorise sa réutilisation. L'idée est donc de passer du contrôle qualité au guide sur la bonne façon de la constituer, mais la normalisation reste l'apanage d'une fraction infime des jeux de données, or la qualification des données au regard des usages s'avère essentielle.

Quitte à se projeter vers l'avenir, pourquoi ne pas imaginer que la donnée Open Data s'inspire des mécanismes de versionnement et de remontées utilisateurs des logiciels OpenSource ? Ceci contribuerait à tracer à la fois les usages et la qualité des données.

Décision / Actions :

- effectuer une revue des services de remontées de la qualité des données via des formulaires de contacts (M. Rajerison)

6. Formation à distance

Par S. Levêque

La Formation à distance sur la qualification des données géographiques s'appuie sur les [fiches méthodologiques](#). La session dure théoriquement 18 h pour l'apprenant réparties

sur trois semaines. Il s'agit d'une FOAD tutorée, mobilisant 2 à 4 tuteurs.

Il n'y a jusqu'ici aucun déploiement du fait de la difficulté à identifier un CVRH volontaire pour cette action. En particulier le CVRH d'Aix n'avait pas donné suite.

La possibilité d'un module de sensibilisation avec le Centre National de la Fonction Publique Territoriale (CNFPT) n'a pas encore été étudiée.

S. Levêque recherche donc comment organiser cette FOAD. Elle pourrait soit être prise en charge et organisée par le CEREMA, soit organisée avec OpenDataFrance.

N. Py demande si l'ENSG pourrait organiser la formation test. Il s'interroge également sur la possibilité pour les IDG de prendre le relais sur ce type de formations.

De son côté OpenDataFrance se laisse d'attendre le CNFPT pour organiser sa propre formation orientée open data. Il suggère d'étendre sa propre formation d'une part à la smart data, et d'autre part à la qualité des données géographiques.

Décision / Actions :

- *Monter une session test de la FOAD pour des collectivités sur la base de 2 scénarios :*

. *prise en charge par le Cerema (S. Levêque avec M. Rajerison)*

. *prise en charge par OpenDataFrance (J-M Bourgogne avec S. Levêque, et F. Boithias)*

7. Atelier : synopsis pratique pour la qualification de données

Par S. Levêque, M. Rajerison et S. Rolle

Il s'agit de construire un synopsis permettant de déterminer, en fonction de la nature d'une donnée, quels critères / quelles mesures sont à utiliser et dans quel ordre.

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 listant les données du SCDL (ou autres données). Il s'agit de 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. Afin d'alimenter la réflexion globale sur la méthode, des expérimentations pratiques ont été réalisées sur les données « Base Adresse locale » et « Equipements collectifs publics ».

La session reprend cette réalisation sous la forme de deux ateliers menés en parallèle : Base adresse locale et Equipements collectifs (en fait : établissements recevant du public), suivis d'une restitution collective.

Les ateliers ont consisté en :

- l'analyse par groupe des tableaux portant priorisation et pondération des mesures de qualité, afin de les finaliser.
- établir les liens avec les mesures du « registre français pour les Métadonnées relatives à la qualité des données géographiques » ;
- expliquer la démarche dans une fiche méthodologique. Exemple : [la note méthodologique expliquant le cas « Adresse »](#).

Les deux tableaux ont été produits :

- [qualification de la donnée « Adresse » et/ou « Base adresse locale »](#)
- [qualification de la donnée « Equipement collectifs \(recevant du public\) »](#).

Il ressort de ce double atelier :

- qu'une fiche méthodologique générale pourrait s'appuyer sur un arbre de décision, ou une carte mentale ;
- qu'il convient de compléter le registre (ou thésaurus) des usages génériques de l'information géographique : cartographie, calcul d'itinéraire, etc., afin de bien placer les priorités et pondération des mesures de qualité en face des usages.

Extrait du CR précédent : le groupe « Adresse » avait identifié des classes d'usages pour l'adresse :

- Administratif : gestion administrative ; certificats d'adressage ; fiscalité (localisation d'un local, déclaration d'une propriété)
 - Défense et secours : défense externe contre les incendies ; secours, etc.
 - « Vie quotidienne » : navigation ; publipostage ; livraison postale
 - Usages gestionnaire et techniques : localisation ; géocodage ; aménagement numérique du territoire, etc.
- Avec d'autres clés d'entrée des usages :
- usages professionnels (lorsque la qualité de service dépend de l'adresse), versus usages non professionnels
 - usages « machine to machine » versus « human to machine »

Décision / Actions :

- S. Levêque et S. Rolle finalisent les tableaux pour les deux cas d'usage « Adresse » et « Equipco ».
- S. Levêque et C. Jamet dressent l'arbre de décision pour le cas d'usage « Adresse »
- M. Rajerison et S. Rolle dressent l'arbre de décision pour le cas d'usage « Equipco »
- J-M. Arzac élabore le thésaurus des usages génériques de l'information géographique (dont une ébauche est disponible dans la [note méthodologique du cas « Adresse »](#))
- N. Py présente la méthodologie générale sous forme de carte mentale (sous Freemind ou Freeplane)

8. Atelier : restitution de la qualification des données géographiques

Les participants ont discuté de la meilleure façon de restituer graphiquement la qualité d'un jeu de données géographiques. Il en est ressorti que les diagrammes polaires offrent une excellente solution à condition :

- de les distinguer suivant l'usage technique attendu de la donnée ;
- et qu'ils portent sur leurs branches les mesures correspondant aux différents critères de la qualité de l'information géographique.

A l'instar du site « Les numériques » le groupe de travail s'oriente vers :

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

recommandés ».

RÉSUMÉ

NOTE DE LA RÉDACTION

★★★★★

NOTE DES UTILISATEURS (2)

★★★★★

VOTRE AVIS

[DÉPOSER UN AVIS >](#)

SOUS-NOTES

ERGONOMIE ET DESIGN	★★★★★	ÉCRAN	★★★★★
PERFORMANCES	★★★★★	PHOTO	★★★★★
AUTONOMIE	★★★★★		

POINTS FORTS

- ✓ Finitions d'excellente facture.
- ✓ Dalle Oled bien calibrée.
- ✓ Autonomie monstre.
- ✓ Prise jack de qualité.
- ✓ Trois modules photo excellents (sauf en faible luminosité).

POINTS FAIBLES

- ✗ Photos en faible luminosité qui ne rivalisent pas avec celles d'un Google Pixel 3a.
- ✗ Pas de slot pour carte microSD.

NOTE DES UTILISATEURS (2)

★★★★★

Note des utilisateurs (2)

Avis utilisateurs (2)

5★

4★

3★

2★

1★

Autonomie	★★★★★
Écran	★★★★★
Ergonomie et design	★★★★★
Performances	★★★★★
Photo	★★★★★

VOIR TOUS LES AVIS

Décision / Actions :

- S. Levêque élabore des maquettes de représentation inspiré de sites web tels que : « Les numériques », « Allo-ciné », « Que-choisir », ...