



Liberté • Égalité • Fraternité

RÉPUBLIQUE FRANÇAISE



Conseil national de
l'information géographique

Commission règles & qualité

**Groupe de Travail Qualité
des Données Géographiques (QuaDoGéo)**

Groupe de Travail animé par :



Direction territoriale Ouest
Département Villes et territoires

**COMPTE-RENDU DE REUNION
(IGN saint-Mandé, salle A571)**

Objet : Réunion #3 du GT QuaDoGéo du 18 septembre 2019.

Ordre du jour :

1. Validation du [précédent compte-rendu](#). Points d'info et d'actu.
2. Travaux de qualification réalisés par l'IGN sur ses bases de données (C.Coulomb)
3. Production de synopsis pratiques pour la qualification de données
 - Base adresse locale
 - Equipco
4. Registre français pour les "Métadonnées relatives à la qualité des données géographiques" (A. Gallais)
5. Restitution des éléments de qualification d'un lot de données - apports du standard Geospatial User Feedback (GUF) (S. Rolle)

Liste des participants

Animateur : Arnauld GALLAIS (CEREMA)

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
Chantal COULOMB	IGN / Contrôle Qualité	x	
Olivier DISSARD	CGDD/DRI/MIG	x	
Loïc HAÏ	OpenDataFrance / Validata		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
Stéphane ROLLE	CRIGE PACA	x	
Pierre VERGEZ	IGN / Mission CNIG		x

Réunion #3 du GT QuaDoGéo du 18 septembre 2019.		
Mise en relecture du compte-rendu	Sans	Arnaud Gallais
Relecture du compte-rendu	Sans	Les participants.
Validation du compte-rendu	08/10/2019	Arnaud Gallais
Prochain rendez-vous le 21 novembre 2019 en présentiel à Lyon (Cerema).		

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

- Le [compte-rendu \(permalien\)](#) du [GT QuaDoGéo](#) du 4 juin est relu et validé.

- **A.Gallais** informe du lancement en juin par [Etalab](#) de [schema.data.gouv.fr \(permalien\)](#). Cette initiative est liée au constat que « *la montée en qualité des données est aujourd'hui l'un des principaux objectifs de l'open data, notamment pour faciliter la réutilisation des données. Les schémas permettent entre autres de valider qu'un jeu de données se conforme à un schéma, de générer de la documentation automatiquement, de générer des jeux de données d'exemple ou de proposer des formulaires de saisie standardisés* ». il s'agit d'une action à rapprocher de la plateforme Validata développée par OpenData France. Loïc Haÿ n'y travaillant plus, il est prévu d'inviter **J-M.Bourgogne** pour une représentation OpenDataFrance au GT QuaDoGéo.

- **A.Gallais** informe qu'une problématique de qualification des données géographiques est apparue au [GT CNIG DDU \(permalien\)](#) du 12 septembre : le validateur GPU ne contrôlant que certains aspects de la cohérence logique (et probablement en laissant de côté certaines contraintes topologiques), 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 société 1Spatial dont les clients de la solution PCI Base ont relevé ces difficultés sera invitée au prochain GT QuaDoGéo.

- Formation à distance sur la qualification des données géographiques (cf. [CR précédent](#) et notamment le retour d'expérimentation de **S.Rolle**). La formation est toujours en période de test et il n'y a pas eu d'avancée sur son déploiement durant la période estivale du fait de la difficulté à identifier un CVRH volontaire pour cette action. Il est prévu de relancer le CVRH d'Aix. **J.Baret** se propose de tester la formation avec quatre personnes du Conseil Départemental du Loiret, **M.Le-Moal** (Axes conseil) qui donne également des formations sur la qualité des données géographiques, ainsi que **J-M.Arsac** (Azimut) pourraient également se montrer intéressés. Le CRIGE PACA recherchera également des stagiaires potentiels. La possibilité d'un module de sensibilisation avec le Centre National de la Fonction Publique Territoriale (CNFPT) n'a pas encore été étudiée.

- **A.Gallais** informe le GT QuaDoGéo de différents points :

- Un cas d'utilisation : un audit de la base de données arrêts du réseau régional Bretagne sera prochainement réalisé. Mené par le Cerema Ouest, cette action portera sur la qualification des cinq bases de données (des quatre départements et de la région), en s'interrogeant notamment sur leur homogénéité. Il s'agira également de conseiller la région sur les caractéristiques d'une base améliorée avec les actions de correction nécessaires. L'audit des bases existantes s'effectuera au regard de la norme ISO 19157 en l'appliquant, a priori pour la première fois, sur des données « General Transit Feed Specification » (GTFS).

- Un sujet de « [projet professionnel pluridisciplinaire](#) » ([permalien](#)) sera proposé à un groupe d'élèves ingénieurs de l'ESGT : « Expression de la qualité dans les géostandards »¹. L'objectif de ce projet est de comparer l'expression de la qualité des données dans différents géostandards, au regard des critères qualité des normes ISO 19115 et ISO 19157, eux-mêmes résumés dans les [fiches méthodologiques du Cerema « Qualifier les données géographiques » \(permalien\)](#). Cette comparaison visera à mettre en exergue les bonnes pratiques pour ensuite élaborer des recommandations pour

¹ Révision 2024 : Ressource perdue.

l'ensemble des géostandards (CNIG, covadis, SDCL, etc.), relatives au domaine de la qualité. [hors réunion : le sujet n'a malheureusement été retenu par aucun groupe d'étudiants...]

- Pour mémoire, **N.Py** a communiqué au GT QuaDoGéo :
 - une consolidation de la norme ISO 19157 au format [XML sur Github](#). ([permalien](#)).
 - la publication « [L'information géospatiale légale](#) » ([permalien](#)) dressant le bilan des productions d'Armelle Verdier pour le compte de l'AFIGEO en 2019, avec des thèmes touchant de près à la qualification et la sécurité juridique des données géographiques, notamment « [Les enjeux juridiques du Géoportail de l'Urbanisme](#) » ([permalien](#)) et la « [Réflexion sur la notion de fiabilité juridique des données géographiques](#) » ([permalien](#)), qui présente la qualité des données géographiques sous l'angle d' « une problématique mal connue du droit ».

Décision / Actions :

- Inviter **J-M. Bourgogne** pour une représentation OpenDataFrance au GT QuaDoGéo (**S. Levêque**)

- inviter la société 1Spatial à témoigner des soucis de qualité des données d'urbanisme publiées sur le GPU (**A. Gallais**)

- ré-impulser l'action de déploiement de la FOAD, et monter une session test pour des collectivités (**S. Levêque** avec **M. Rajerison**, **S. Rolle**, **J. Baret**)

- étudier / exploiter la consolidation de la norme ISO 19157 au format XML sur Github.

2. Travaux de qualification réalisés par l'IGN sur ses bases de données

Voir [la présentation](#) ([permalien](#)) de **C. Coulomb** (IGN / Direction des opérations et des territoires)

L'IGN réalise de longue date des contrôles qualités sur de multiples produits issus de ses chaînes de production : BDUNI, BD OCSGE, BD Forêt, produits issus d'images satellitaires, séries cartographiques, etc...

Pratiquement tous les contrôles comprennent une phase de contrôle terrain (avec des outils de positionnement précis) sur des échantillons définis par « échantillonnage aléatoire stratifié » c'est-à-dire constitués dans des strates homogènes, telles que : milieu urbain, péri-urbain, rural, etc.

Les contrôles portent particulièrement sur les critères d'exhaustivité, de précision géométrique et de précision thématique.

L'analyse des contrôles bénéficie désormais des compétences d'un statisticien utilisant le logiciel R.

Les résultats des contrôles sont évalués à l'aune des spécifications avec un code couleur : rouge si la production ne répond pas aux exigences, vert s'il n'y a pas d'anomalies, et noir s'il n'y a pas d'exigence particulière.

Des représentations plus détaillées et assez conviviales (par exemple des curseurs sur une échelle de rouge à vert) sont également réalisées mais aujourd'hui réservées à l'utilisation interne à l'IGN. Le GT QuaDoGéo y verrait un intérêt pour une diffusion en externe, accompagnés si nécessaires des données brutes issues du contrôle qualité.

C.Coulomb signale que les contrôles effectués engendrent systématiquement des corrections, voire améliorations de spécifications de saisie pour les produits contrôlés, dans le cadre d'un cycle d'amélioration continue.

Afin d'aider les utilisateurs à apprécier objectivement la qualité des produits, l'unité de contrôle qualité élabore des fiches d'aide pour accompagner les résultats, mais ces fiches ne sont pas encore finalisées.

L'IGN souhaite s'impliquer dans leur finalisation, ainsi que dans la définition d'un processus de contrôle pour le PCRS.

Les débats qui ont suivi la présentation ont porté sur l'intérêt de différencier entre la diffusion des évaluations de qualité pour l'interne et l'externe, et sur l'importance d'une démarche participative avec les utilisateurs de produits IGN dans le style « design de service » pour bien identifier leurs souhaits en termes de restitution des éléments de qualification des données : quels éléments ? sous quelle forme ? avec quel niveau de détail ? On pourrait probablement envisager des éléments de restitution correspondant à différents niveaux de lecture : basique, intermédiaire, avancé..

J.Baret est d'avis qu'il serait également intéressant de diffuser ces éléments suivant des grands domaines d'usage des données (cartographie, navigation, statistique, etc.) avec un côté dynamique permettant à l'utilisateur d'identifier ses usages, de choisir ses indicateurs et de produire un rapport qualité dynamique au sens où il serait directement en rapport avec ces usages.

3.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 listant 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.*

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

- 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**

Base Adresse Locale (BAL)²

Voir [ici la présentation \(permalien\)](#) de **S.Levêque**

Deux séances de travail ont réuni **C.Jamet, N.Py, A.Gallais** et **S.Lévêque** pour poursuivre la réflexion sur le [synopsis de qualification des données géographiques en prenant le cas de la base adresse locale \(BAL\)](#).

² Révision 2024 : la gouvernance et le schéma des BAL ont évolué depuis cette réunion ; les permaliens sont plus récents et ne refléteront que partiellement le sujet tel qu'il pouvait être travaillé à l'époque.

L'équipe a recentré ses réflexions sur le [format d'échange \(BAL\)](#) ([permalien du format en v1.1](#)) défini dans le [socle commun de données locales](#) (SCDL, [permalien](#)).

Elle a constaté que quand on cherche à qualifier on est très rapidement tenté de faire des retours et propositions pour l'amélioration du produit.

L'exercice sur la BAL a porté sur : l'évaluation de l'importance intrinsèque des critères de qualité ; l'identification des cas d'usage ; l'ordonnancement à retenir pour la qualification du type de données « adresse » ; la détermination de la dépendance aux données externes ; en y ajoutant des commentaires et/ou points de vigilance.

L'utilité d'un thésaurus sur les usages génériques de l'information géographiques est rapidement apparue. Ce thésaurus serait à identifier s'il existe, ou bien à créer.

Exemple : le sous-groupe a 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 »

A l'instar du sous-groupe étudiant le cas des équipements collectifs, celui sur la BAL préconise une représentation graphique sous forme de diagramme polaire pour synthétiser la notation et l'indice de confiance de la donnée en fonction des usages.

A. Gallais rappelle que l'expérience doit à terme déboucher sur une note méthodologique décrivant comment l'utilisateur doit opérer pour qualifier les données (*cf. attendus cités dans le §5 du [mandat](#) du GT CNIG QuaDoGéo*).

Equipements collectifs (Equipco)

Pas d'avancée significative depuis la dernière réunion.

Décision / Actions :

- Poursuivre les travaux, en y associant si possible OpenData France
- 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 : envoyer l'état des réflexions pour avis et commentaire, et réaliser un test de qualification de données.
- Equipco : réaliser un test de qualification des données de Datasud
- BAL et Equipco : rédiger et partager le projet de note méthodologique
- Démarrer la réflexion sur « Aménagements cyclables » et « Admin express »

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

Le sujet consiste à guider et faciliter la saisie et la transmission d'informations sur la qualité de la donnée en créant des rubriques dédiées au sein des métadonnées. Il s'agit donc d'améliorer la saisie des critères qualité dans les métadonnées ISO / Inspirer 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 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)

Ce registre pourra 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](#) ([permalien](#)) peut être utilisé pour le constituer.

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

A.Gallais présente le [projet de registre français des métadonnées de qualité des données géographiques](#) ([permalien](#)), qui a préalablement été visé par **M.Lambois** de la cellule normalisation de l'IGN.

Décision / Actions :

- intégrer le critère actualité / fraîcheur des données dans le registre (**A.Gallais**, **M.Lambois**)
- étudier le cas des données images en prenant appui sur des spécifications de métadonnées de produit raster (**A.Gallais**)
- implémenter le registre en xml
- étudier comment son intégration pourrait s'effectuer dans [Géonetwork](#)

5. Geospatial User Feedback (GUF)

Voir [la présentation](#) ([permalien](#)) de **S.Rolle**

Cet outil est conforme aux standards de l'OGC. Le principe est de permettre (ou de demander) aux utilisateurs de qualifier les données en fonction de l'utilisation qu'ils en ont faite, il s'agit donc d'une « qualification par les pairs ».

S.Rolle est d'avis que le projet est intéressant mais encore neuf (2018) et instable. Surtout, l'interface s'avère actuellement complexe et peu ergonomique dans la mesure où elle semble replacer les utilisateurs dans une situation de type « saisie de métadonnées »...

L'outil Ripart de l'IGN est également cité, mais il est davantage conçu pour saisir des alertes d'évolutions sur le terrain que pour effectuer de véritables retours sur la qualité des données.

Décision / Actions :

- *Etudier le modèle plus en profondeur, ainsi que les briques qualité des données (M. Rajerison)*
- *Tester le [portail de démo nimmbus opengis](#)³ (M. Rajerison)*
- *Evaluer l'opportunité d'implémentation sur une plateforme de diffusion de données.*

Prochaine réunion le 21 novembre 2019 en présentiel à Lyon (Cerema).

³ Révision 2024 : ressource perdue.