

## Participants :

Nom Prénom	Organisme	Présent	Excusé
Geoffrey ALDEBERT	Etalab - data.gouv.fr		x
Bernard ALLOUCHE	Cerema / DG / supervision des données		x
Jean-Marie ARSAC	AZIMUT	x	
Mathieu BECKER	ISOGEO	x	
Jérôme BOUTET	Idéo Ternum Bourgogne-Franche-Comté	x	
Arnaud BRAUN	IGN/DSI/SDM/Géomatique et Carto.	x	
Pierre CAMILERI	Multi / OpenDataFrance / Validata	x	
Benjamin CHARTIER	Consultant - anim. GT Open Data Afigéo		x
Clémence COTTEN	IFREMER	x	
Chantal COULOMB	IGN / Contrôle Qualité	x	
Adeline COUPÉ	IGN / Géoplateforme	x	
Laurène DEBRAY	Office international de l'eau / Sandre		x
Héloïse DESCHAMPS	GIP Atgeri	x	
Olivier DISSARD	MTE/CGDD Admin données algos codes		x
Arnaud GALLAIS	Cerema Ouest	x	
Thomas GRATIER	Etalab - data.gouv.fr		x
Guillaume GRECH	UMS Patrinat, OFB - MNHN		x
Marie LAMBOIS	IGN / Cellule normalisation	x	
Sebastien LAUNAY	IGN / Dépt Contrôles Qualités	x	
Mathieu Le Moal	Axes Conseil	x	
Stéphane MEVEL-VIANNAY	Région Bretagne / GéoBretagne	x	
Steven PIEL	OFB / <a href="#">Service des référentiels marins</a>		x
Nicolas PY	IGN Centre Est	x	
Mathieu RAJERISON	Cerema Méditerranée	x	
Gessica REYNAUD	Géomap - Imagis		x
Stéphane ROLLE	CRIGE PACA	x	
Cédric LEPICIER	stagiaire	x	
Benoît SEGALA	Consultant - Cabinet d'études	x	
Pierre VERGEZ	IGN / Mission CNIG		x
David VIGLIETTI	Office international de l'eau / Sandre	x	

## Ordre du jour :

- Validation du [précédent compte-rendu](#). Points d'info et d'actu
- Contrôle qualité de la production OCSGE nouvelle génération (*A. Braun*)
- Avancée de la normalisation - Métadonnées qualité dans DCAT (*M. Lambois*)
- Classification des usages de l'information géographique (*J-M. Arsac et S. Rolle*)
- Projet Validata JSON pour OpenDataFrance (*J. Richer*)
- « Chaîne de qualité générique » basée sur FME (*S. Rolle*)
- Point divers

Les documents relatifs à cette réunion sont [disponibles ici](#).

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

### Revue des actions du précédent compte-rendu :

#### Actions présentées en séance :

- *Consolider la nomenclature des usages (J-M. Arzac, S. Rolle, J. Boutet)*

#### Actions à lancer ou poursuivre :

- *Partager des modèles de rapport de contrôle qualité (tous)*  
 - *Suivi du dispositif GUF Géonetwork v4 (M. Lambois, C. Rabevolo)*  
 - *Rencontre bilatérale Sandre - IGN pour approfondir les questions relatives aux processus de contrôles qualité et leur restitution aux producteurs et aux utilisateurs (N.Py)*  
 => cette action est désormais déléguée au produit BD Topage. Le GT QuaDoGéo maintiendra une veille sur ce sujet.

### Point d'infos & actus :

- S. Launay a transmis des [exemples de contrôles qualité](#) concernant plusieurs produits de l'IGN : BD Adresse, BD Uni, RGE Alti, BD Ortho... Les membres du GT QuaDoGéo compléteront cette documentation de leurs propres exemples pour constituer une "bibliothèque de contrôles qualité" utilisables comme modèles.

C. Coulomb indique que l'IGN établit des modèles de contrôle qualité en vue de les diffuser d'ici fin 2022 via [Géoservices](#).

- Pour permettre d'identifier plus facilement les jeux de données de qualité et accompagner la documentation des données, [data.gouv.fr](#) a mis en place un score de qualité des métadonnées, à découvrir sur la page des (presque 43.000) [jeux de données](#).

[data.gouv.fr](#) attend les retours de la communauté des utilisateurs.

The screenshot displays the 'data.gouv.fr' interface. A central dataset, 'Répertoire national des élus RNE' by the 'Ministère de l'Intérieur et des Outre-Mer', is highlighted. It has a quality score of 100% (indicated by a green bar) and was last updated on September 13, 2022. A dropdown menu for 'Qualité des métadonnées' is open, showing the following criteria: 'Description des données renseignée' (checked), 'Licence renseignée' (checked), 'Formats de fichiers standards' (checked), and 'Couverture spatiale renseignée' (checked). Other criteria like 'Documentation des ressources manquante', 'Fréquence de mise à jour non respectée', and 'Couverture temporelle non renseignée' are marked with warning icons. To the right, statistics for this dataset show 17 réutilisations and 51 favoris. Below it, another dataset 'Santé publique France' is visible with 144 réutilisations and 63 favoris. At the bottom, the 'Insee' dataset is partially visible with 34 réutilisations and 48 favoris.

- La formation "[Qualité des données géographiques](#)" organisée par le CRIGE (S. Rolle) et le CEREMA (M. Rajerison) se poursuit dans un format court ([trois sessions de 1h fin 2022](#), et potentiellement 2023) en partenariat avec le CD30 et OpenIG et davantage orienté vers l'utilisation des outils pour mettre en œuvre les contrôles qualité (ex : Postgre/PostGIS, FME, etc). S. Rolle rappelle que ces sessions sont gratuites et ouvertes au niveau national.

- A. Gallais informe qu'après quatre années d'animation du GT CNIG QuaDoGéo il souhaite se consacrer à d'autres chantiers sur les données. Ces quatre années, (initialement la reprise du flambeau de l'action de G. Troispoux du Certu), ont tracé un parcours jalonné d'objectifs opérationnels avec des avancées (consolidation des fiches méthodologiques, registre national des mesures qualité, formations, nomenclature, outil de restitution graphique, etc.) A. Gallais remercie l'ensemble des intervenants de cette production et des nombreux retours d'expériences qui ont marqué ce parcours.

Les volontaires pour reprendre l'animation du GT QuaDoGéo se manifesteront auprès de M. Lambois, présidente de la [Commission "Règles et Qualité"](#) du CNIG.

## 2. Contrôle qualité de la production OCSGE nouvelle génération

*Par A. Braun. La présentation est [disponible ici](#).*

L'Occupation du Sol Grande Échelle (OCS GE) décrit tout point du territoire par des polygones portant deux informations : la couverture (ce que c'est, ce que l'on voit), et l'usage (à quoi cela sert), selon une nomenclature hiérarchique, conforme aux [préconisations du CNIG](#). Ces prescriptions comprennent des seuils de dimensions minimales.

La production s'est récemment accélérée dans le contexte de maîtrise de l'artificialisation des sols. La DGALN a commandé une production massive et rapide de données d'occupation du sol à grande échelle couvrant tout le territoire. Un saut technologique a ainsi été opéré par l'IGN en introduisant des techniques d'intelligence artificielle (deep learning) dans la chaîne de production.

Un soin particulier est apporté aux questions d'évaluation de la qualité tout au long de la chaîne, pour sécuriser la production :

- En amont de l'apprentissage IA, des annotations sont validées par contrôle qualité. Les imagerie de 512x512 pixels servant à l'apprentissage du modèle doivent être annotées avec la meilleure certitude possible. Ces annotations sont contrôlées par échantillonnage et vérification. En cas de désaccord de l'opérateur de recette avec la classe annotée proposée par le producteur, le second regard d'un autre opérateur de recette permet de trancher, cela permet de lisser la variabilité.
- Le modèle IA est évalué par des méthodes éliminant les biais liés à la répartition des annotations. On produit ainsi les métriques classiques de la littérature : précision, rappel, F-Score, IoU, précision globale (overall accuracy).
- Le produit final OCS GE est quant à lui vérifié par des procédures classiques de contrôle qualité mis en œuvre via des scripts et des contrôles interactifs sous ArcGIS : validation de la cohérence logique, estimation de la précision thématique par échantillonnage et calcul statistique, édition de matrice de confusion.
- Enfin, sur certains départements, un contrôle qualité complémentaire est effectué « à froid » pour apporter des informations complémentaires. Ce contrôle de la qualité s'effectue suivant la méthode de contrôle maintenant bien rodée de la production OCS GE classique.

La couche des zones construites est un co-produit de l'OCS GE obtenu par agglomération de classes . Elle permet de cerner les zones à enjeu. Elle est également contrôlée.

### 3. Avancée de la normalisation - Métadonnées qualité dans DCAT

Par M. Lambois. Les présentations sont disponibles ici : [normalisation - MD qualité DCAT](#)

#### 3.1 Avancée de la normalisation

Dans le Data Quality DWG des dernières plénières de l'OGC, une présentation de Heiko Figgemeier de l'Université de Dresde portait sur un "géo-dashboard" : tableau de bord de la qualité. Partant du constat qu'il est nécessaire de partager différents éléments de qualité, il a été conçu en plusieurs parties afin de présenter différents aspects de la qualité : la source des données (via des graphes de provenance), leur qualité statistique, leur qualité en fonction des zones ainsi que différents éléments de métadonnées. Ce projet est disponible [sur ce Github](#) et il confirme l'importance de la connaissance de la généalogie des données, qui sont de plus en plus ré-utilisées.

#### 3.2 Métadonnées qualité dans DCAT

Le [GT Métadonnées](#) du CNIG travaille actuellement à une mise en correspondance des éléments de [métadonnées Inspire du guide national](#) avec les éléments de métadonnées de type DCAT (et/ou geoDCAT).

Les travaux sont portés par [ce Github](#), on l'on trouve une ébauche de [correspondance Inspire => DCAT](#). Il s'agit de commenter les correspondances via [des issues](#).

Le [mapping actuel des éléments de qualité](#) (issu des travaux européens) est un peu pauvre et certains critères de qualité (type cohérence topologique) n'ont pas de correspondance. Pour pallier à ce problème une option serait d'intégrer les éléments de métadonnées du vocabulaire Data Quality du W3C.

La prochaine réunion du GT CNIG Métadonnées se tiendra le 21 octobre à 14h.

#### [Questions / débats](#)

- Il convient de coordonner GT Métadonnée et GT QuaDoGéo via un sous-groupe "pivot" du GT QuaDoGéo

- A. Gallais constate que la correspondance avec DCAT prend en compte 4 éléments Généalogie ; Résolution spatiale ; Cohérence topologique et Conformité des données mais ne doit-on pas prendre en compte toutes les mesures de qualité référencées dans le [Registre français des métadonnées relatives à la qualité des données géographiques](#) ? Une [issue](#) a été créée à ce sujet.

#### [Décision / Actions :](#)

- *Coordination des aspects qualité entre GT Métadonnées et GT QuaDoGéo (J. Boutet à partir de 2023, éventuellement accompagné de N. Py)*

### 4. Classification des usages de l'info géographique

Par J-M Arsac, S. Rolle, J. Boutet.

La grille de lecture est [disponible ici](#).

*Le groupe de travail a remodelé la nomenclature des usages établies en 2020 par J-M. Arsac dans le cadre de la [méthodologie pratique pour qualifier des données](#).*

*Cette nomenclature constitue l'élément indispensable à la qualification de la donnée par rapport à ses utilisations potentielles (qualité externe).*

Un [document collaboratif](#) (framapad) a été mis en ligne que la communauté des utilisateurs pouvait abonder avec les différents usages identifiés.

Le groupe de travail a constaté la nécessité d'adopter deux entrées : l'une par la thématique concernée, l'autre par l'usage.

Des nomenclatures existent mais, par exemple, le catalogue INSPIRE apparaît trop détaillé pour être générique. La même remarque peut être faite pour la [nomenclature des usages](#) exposée par l'IGN sur son site Géoservices.

Une difficulté de l'exercice semble d'établir le juste compromis entre une trop grande généralité des usages (qui seraient passe-partout) et des usages trop précis (pas assez généraux).

Le GT QuaDoGéo s'est donc prononcé pour la sélection d'une courte liste d'actions (un verbe à l'infinitif) correspondants à une dizaine d'usages génériques.

La nomenclature a été finalisée. Elle a été réduite à 8 usages fondamentaux de l'information géographique.

**7 usages génériques** : [Visualiser / cartographier] ; [Inventorier / recenser] ; [Localiser / situer] ; [Suivre / observer] ; [Analyser] ; [Gérer] ; [Planifier]

et **1 usage transversal** : [servir de donnée référentielle et/ou de donnée pivot]

Le sous-groupe a établi le [tableau de correspondance](#) entre la [nomenclature des usages thématiques](#) exposée par l'IGN sur son site Géoservices et ces 8 usages génériques. Ce principe pourra être adopté pour toute nouvelle thématique.

#### [Questions / débats](#)

La nomenclature des usages est validée par le GT CNIG QuaDoGéo et sera soumise à commentaires et/ou contributions sur un Github. Celui du GT QuaDoGéo est créé à cette occasion.

#### [Décision / Actions :](#)

- Création du [Github QuaDoGéo](#) et invitation des participants (M. Lambois) (c'est fait !)
- Publication de la nomenclature (J-M Arsac)
- Relayer l'info via Géorezo

## 5. **Projet Validata JSON pour OpenDataFrance**

Par P. Camileri. La présentation est [disponible ici](#).

Dans la continuité de [Validata](#) (renommé pour l'occasion Validata Table), Validata JSON est une API de validation de données ouvertes au format JSON ou GeoJSON, avec l'aide d'un schéma au format "JSON schema", suivant l'exemple des [données d'aménagement cyclables](#) avec le [schéma correspondant](#).

Validata JSON encapsule le moteur python de l'open knowledge foundation et répond à plusieurs besoins fonctionnels. Il offre premièrement un service sous forme d'une API asynchrone, qui permet de lancer de longs traitements sans attendre le résultat, qui pourra être récupéré ultérieurement. Le service est temporairement hébergé pour être testé avec [transport.data.gouv](#), la documentation est accessible [ici](#).

Ensuite, il propose un rapport d'erreur en français et abordable, y compris pour les utilisateurs ayant moins de compétences techniques. Le rapport de validation consolide des erreurs similaires survenues à des index différents dans un tableau, pour éviter de présenter la même erreur plusieurs fois. La présentation de l'erreur est également nettement plus détaillée que dans les validateurs classiques.

Enfin, il était important que le validateur puisse traiter des fichiers GeoJSON de gros volumes, plusieurs centaines de Mo, à l'instar du fichier France entière des aménagements cyclables. Pour ce faire, les gros fichiers sont découpés en plusieurs petits fichiers, qui sont mis dans une file d'attente ("task queue"), et traités en parallèle par des validateurs ("workers"). Enfin, le résultat est ré-agrégé en un rapport d'erreur unique. La file d'attente permet également d'absorber la charge de plusieurs demandes de validation simultanées.

#### Questions / débats

Cette solution intéressante et, pour une fois, conviviale pour l'utilisateur. Elle est actuellement en phase de tests sur transport.data.gouv.fr. L'outil sera déployé début 2023.

Entre autres perspectives, les petits fichiers pourraient être priorisés par rapport aux fichiers correspondant à une emprise nationale.

La solution fonctionne a priori avec n'importe quel modèle de données conforme à "Table schema" préconisé par schema.data.gouv.fr.

N. Py pose une alerte quant à la précision (pas mieux que métrique) des données géographiques référencées dans le système WGS84 EPSG4326, a priori seul système de référence supporté par le format GeoJSON. M. Lambois indique que le support d'autres systèmes de référence, par le format GeoJSON [est à l'étude](#).

## **6. Chaîne de qualité générique, basée sur FME**

*Par C. Lepicier et S. Rolle. La présentation est [disponible ici](#).*

Dans le souci de diffusion de données optimisées et qualifiées, le CRIGE PACAmène depuis plusieurs années des démarches autour de la qualité des données géographiques, et souhaite passer à la phase opérationnelle.

En lien avec la norme ISO 19157 et les critères qualité, le CRIGE PACA a proposé un stage à Cédric Lepicier en lui demandant de mettre en place une chaîne de traitement générique de qualification de données géographiques.

Cette chaîne, mise en place avec le logiciel FME, a pour objectif de contrôler des données géographiques, sur une emprise spatiale de référence, selon deux indicateurs/sous-indicateurs de la norme : la cohérence de format et la cohérence topologique.

Le processus permet également d'envoyer automatiquement des rapports d'état au « testeur ». Il n'est pas question ici de rapport de conformité mais bien de rapport d'état au regard de critères qualité. Les autres critères seront traités ultérieurement.

L'objectif final est d'informer de la qualité intrinsèque d'un lot de données, sans corriger les erreurs rencontrées. Ce prototype fonctionnel (en mode "poste personnel" pour le moment) a vocation à être déployé avec une version serveur de FME en tant que service opérationnel offert sur la plateforme du CRIGE .

#### Questions / débats

A. Gallais suggère de partager cette solution au sein de la communauté des utilisateurs de FME.

J. Boutet demande s'il sera possible de mettre en place des partenariats avec d'autres plateformes pour la formation. S. Rolle répond positivement et indique qu'une discussion à ce sujet est envisageable au sein du réseau des CRIGE.

## 7. Points divers

M. Lambois indique qu'une fiche synthétique sur la qualité du PCRS s'avère nécessaire. Ce sujet sera mis à l'ordre du jour de la prochaine réunion, en coordination avec le GT PCRS.

- A ce stade, la production effective ou en cours du [GT CNIG QuaDoGéo](#) comprend :
- Les fiches méthodologiques du Cerema : sur le [site du Cerema](#) ou en accès individuels : [Introduction](#) - [généralités](#) - [contexte du contrôle qualité](#) - [éléments statistiques](#) - [méthodes d'échantillonnage](#) - [mode de représentation](#) - [cohérence logique](#) - [exhaustivité](#) - [précision thématique](#) - [précision de position](#) - [qualité temporelle](#) ;
  - Les [webinaires de formation à la qualité](#) par le CRIGE PACA et CEREMA
  - Le [registre des mesures de qualité des données géographiques](#) et le [registre en ligne](#) sur le Géocatalogue ;
  - 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 ébauche de 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.
  - Le prototype d'un [outil de restitution graphique de la qualification](#) des données géographiques suivant la [maquette](#).
  - la Nomenclature des usages génériques de l'information géographique
  - (*en cours*) [Modèle de parties qualité et métadonnées des géostandards](#)
  - Le fil de discussion Géorezo : "[Qualité des données géographiques](#)"

