

## Compte-rendu de réunion du 7 février 2020

### Participants :

Nom Prénom	Organisme	Présent	Excusé
Bernard ALLOUCHE	CEREMA / supervision des données	x	
Jean-Marie ARSAC	AZIMUT		x
Jean-Marie BOURGOGNE	OpenDataFrance / Validata		x
Jérôme BOUTET	Idéo BFC - GéoBourgogne	audio	
Romain BUCHAUT	CRIGE PACA		x
Olivier DISSARD	CGDD/DRI/MIG		x
Clément JAMET	Métropole de Lyon		x
Arnauld GALLAIS	Cerema Ouest	x	
Mathieu Le Moal	Axes Conseil	x	
Stéphane GARCIA	IGN / Cellule normalisation	x	
Stéphane LEVEQUE	Cerema DTec TV	x	
Jocelyne MARC	IGN / Mission Qualité		x
Nicolas PY	IGN Centre Est	audio	
Clémence RABEVOLO	IFREMER	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.
  - FOAD QuaDoGéo (S. Levêque)
- Qualification de la donnée du SI sur le milieu marin (C. Rabevolo)
- Qualification des réseaux EP/EU (M. Le-Moal)
- Registre français pour les métadonnées relatives à la qualité des données géographiques (A. Gallais)
- Production de synopsis pratiques pour la qualification de données
  - Carte mentale (N. Py)
  - Base adresse locale (S. Levêque)
  - Equipco (S. Rolle)
- Restitution de la qualification des données (S. Levêque)
- Point divers

Prochaine réunion : **5 juin 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 précédent [GT QuaDoGéo](#) est relu et validé. La plupart des actions prévues ont progressé ou été réalisées.

- [Retours utilisateurs quant à l'utilisation des données](#)

M. Rajerison nous a fait part (mail) d'une [discussion sur les retours utilisateurs quant à l'utilisation des données](#).

- [FOAD QuaDoGéo \(par S. Levêque, voir la présentation ici\)](#)

*La Formation à distance sur la qualification des données géographiques s'appuie sur les [fiches méthodologiques](#). La session dure théoriquement 18h pour l'apprenant réparties sur trois semaines. Il s'agit d'une FOAD tutorée, mobilisant 2 à 4 tuteurs.*

*Il n'y a eu jusqu'ici aucun déploiement du fait de la difficulté à identifier un CVRH volontaire pour cette action. S. Levêque recherche donc comment organiser cette FOAD qui pourrait soit être prise en charge et organisée par le CEREMA, soit organisée avec OpenDataFrance.*

L'organisation de la FOAD avec le CMVRH n'est plus d'actualité. Par contre, un partenariat entre Cerema et OpenDataFrance met en place une série de webinaires de 1h au sujet de « *La donnée dans les territoires* » avec un module « *Qualité des données, notamment géographiques* ».

La session de FOAD complète pourrait être mise en place au printemps (recherche de tuteurs, stagiaires, dates et aspects financiers) dans l'objectif de monter une session test en septembre 2020.

*Décision / Actions :*

*- Monter une session test de la FOAD pour des collectivités (S. Levêque avec M. Rajerison)*

## 2. Qualification de la donnée du SI sur le milieu marin

*Par C. Rabevolo. [La présentation est ici](#).*

**Le Système d'Information sur le Milieu Marin (SIMM)** s'inscrit dans le système d'information de l'État. Il est l'un des trois systèmes fédérateurs au niveau national, issus de la Loi pour la reconquête de la biodiversité (8 août 2016), avec le Système d'Information sur l'Eau (SIE) et le Système d'Information sur la Biodiversité (SIB).

Le SIMM poursuit les objectifs définis dans sont [mandat](#) :

- faciliter le partage et la diffusion des données sur le milieu marin ;
- constituer le point d'accès centralisé à des informations fiables, tenues à jour et facilement compréhensibles, relatives au milieu marin ;
- jouer un rôle essentiel dans la mise en œuvre des politiques publique (Directive Cadre « Stratégie pour le Milieu Marin » (DCSMM), Direction Cadre « Planification de l'Espace Maritime » (DCPEM) etc.).

**Le Service d'Administration des Référentiels (SAR)** assure l'élaboration des éléments du référentiel technique et met en place les moyens techniques permettant de proposer un langage commun. En offrant une vision unique du référentiel de données au sein du SIMM, le SAR facilite l'interopérabilité entre les banques de données et le SIMM, mais

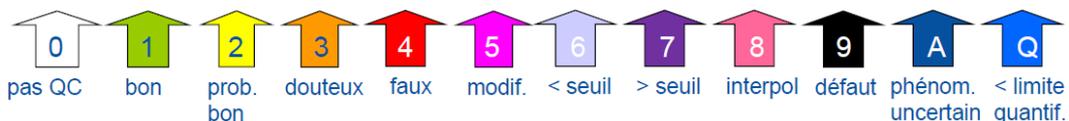
aussi avec le SIE, le SIB et les SI internationaux.

**Le référentiel de qualification de la donnée** permettra d'évaluer les données (référentiels, jeux de données, résultats...) du SIMM de manière standardisée (ex : bonne, douteuse, incomplète, etc). En effet, toutes les données du SIMM sont publiques, y compris des données obsolètes ou invalides, et doivent être diffusées aux autres SI, aux gestionnaires du milieu marin, à la communauté scientifique et au grand public. La connaissance du niveau de qualité est donc indispensable pour que chaque type d'utilisateur sache ce qu'il peut faire (ou non) des données.

Toutes les banques de données du SIMM sont donc concernées par le référentiel de qualification de la donnée essentiel pour assurer l'interopérabilité des données marines dans le cadre du SIMM, mais aussi entre les banques elles-mêmes.

Ce référentiel étant complexe à mettre en place, un brainstorming a été organisé le 30 janvier à Brest pour réfléchir aux moyens de l'adapter aux pratiques des producteurs de données tout en respectant les besoins des utilisateurs. En réponse aux nombreux questionnements soulevés (cf. [présentation](#), diapo 14) le SAR a décidé de :

- s'informer la **norme ISO 19157** et sur les différents retours d'expériences ;
- rédiger les **principes généraux** de la qualification des données du SIMM (par exemple : la qualification doit être effectuée par le producteur, et confrontée à l'approche utilisateur)
- recommander dans un premier temps l'utilisation de **niveaux de qualité** de type SeaDataNet :



**Décision / Actions :**

- *L'analogie des problématiques amènent logiquement le SAR et le GT CNIG QuaDoGéo à la volonté de coopérer sur les méthodes de qualification des données.*

### 3. Qualification des réseaux EP/EU

Par M. Le-Moal. [La présentation est ici.](#)

Les modalités de représentation des résultats d'analyse qualité de données notamment appliquées à des données de réseaux (assainissement et eau potable) ont fait l'objet d'un [article dans Techniques Sciences Méthodes \(TSM\)](#) : « Axes Conseil a élaboré un procédé de contrôle des données SIG rapide et accessible à des acteurs métier de l'eau et de l'assainissement. Plutôt qu'une lourde approche de modélisation a priori, le principe est de générer un ensemble d'indicateurs explicites facilement exploitables a posteriori par les acteurs du métier. »

Deux lots de données sont pris en exemple : les données d'un syndicat des pays de Loire comprenant une dizaine de tables, et celles d'un EPCI de Nouvelle Aquitaine représentées par une quinzaine de bases de données pour un total d'environ 800 tables.

L'analyse de données est automatisée et repose sur les quatre critères : **ATT** (aspect attributaire), **STR** (structure géométrie intrinsèque de l'objet), **REL** (relation spatiale entre les objets) et **RES** (caractéristique réseau, s'appliquant aux données réseaux orientées ou non orientés).

A travers ces exemples, M. Le-Moal différencie la qualification objective liées à l'inventaire par objet et à la synthèse par table (diapos 8 à 12), d'une qualification plus subjective de

l'ensemble de la base de données suivant les critères de la norme ISO 19157. Il souligne que le rapport de qualification doit présenter en premier lieu les aspects généraux avec une part de subjectivité, avant de rentrer dans le détail par tables et par objets suivant les critères ATT, STR, REL et RES.

Concernant le volet réseau (RES), la méthode s'intéresse davantage à la typologie de nœuds de connexion du réseau plutôt qu'à leur positionnement absolu (xy), le plus souvent non défini du fait que l'on ne dispose pas de référentiel géométrique de référence. Les livrables comprennent les données générées par le contrôle et le rapport de qualification. La restitution du niveau de qualification général s'effectue via un de tableau de bord notant chaque critère suivant une échelle de 1 à 5. Au vu de ces éléments de qualification il appartient ensuite au commanditaire (producteur et/ou utilisateur) de définir si le niveau de qualité est compatible avec ses besoins.

Suite aux analyses réalisées, la base de données a qualité variant de bon à excellent suivant les critères (ATT, STR, REL, RES).

Barème	1	2	3	4	5
	Médiocre	Faible	Moyen	bon	Excellent

Critère	Données CCF
Attribut (ATT)	Bon
	<i>Les taux de remplissages sont de bon niveaux</i>
Structure (STR)	Excellent
	<i>La structuration géométrique est de bon niveau. Quelques singularités restent ponctuellement observées.</i>
Relation (REL)	Bon
	<i>Le volumes d'intersection d'objet CANALISATIONS peut se justifier par le modèle de données.</i>
Réseau (RES)	Bon
	<i>La typologie des nœuds et le positionnement des ponctuels sont très satisfaisants. Le positionnement des branchements est plus hétérogène.</i>
Total	Bon

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

Par A. Gallais. [La présentation est ici.](#)

La saisie et la transmission d'informations sur la qualité de la donnée peut s'appuyer sur des rubriques de métadonnées ISO / Inspire spécifiques. Il s'agit d'améliorer la saisie des critères qualité dans les métadonnées 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 des métadonnées de qualité des données géographiques](#).

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

Après avoir initialement pris la forme d'un [document pdf](#), Le registre est désormais

implémenté sur le « [Système de publication de registres](#) » du [Géocatalogue](#) à l'adresse : <https://data.geocatalogue.fr/ncl/mesuresQuaDoGeo>

Chaque mesure qualité est désormais référéncée par un URL, par exemple : le [Taux d'exhaustivité](#), la [Classe de précision au sens de l'arrêté du 16 septembre 2003](#), etc.

Le registre présente actuellement 26 mesures, issues :

- du sous-ensemble de la norme ISO19157 constitué des mesures retenues dans les [fiches méthodologiques du Cerema](#) ;
- de 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)

Les outils de qualification des données (tels que : Validata, validateur GPU, etc...) pourront désormais se référer au registre qui référence par un URL chaque mesure de la qualité.

Le BRGM garde la main sur la publication et l'actualisation du dispositif, aucune actualisation directe n'est possible sans transmission d'un fichier .csv compatible avec le « template » fourni par le BRGM.

Le groupe de travail devra s'attacher à définir les consignes de saisie de métadonnées spécifiques à la qualité (CSMD-Q)

#### Décision / Actions :

- *Soumettre le registre à la validation de la [Commission Règles de mise en œuvre afin de passer de l'état « expérimental » à « stable »](#).*
- *Publication du registre dans ses deux formes pdf et html sur la page de [ressources du GT CNIG QuaDoGéo](#)*
- *Etudier avec le BRGM les pistes d'améliorations du dispositif : possibilité d'hyperlien au sein des entrées du registre ; affichage suivant l'ordre des critères plutôt que le nom des mesures ; affichage suivant l'ordre des items de la norme au sein de chaque entrée du registre.*

## **5. Synopsis pratique pour la qualification de données**

*Par N. Py, S. Levêque 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.*

*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 sont réalisées sur les données « Base Adresse locale » et « Equipements collectifs publics ». La démarche est expliquée dans une fiche méthodologique. Exemple : [la note d'accompagnement expliquant le cas « Adresse »](#).*

Afin d'étayer une fiche méthodologique générale N. Py a réalisé sous Freeplane [une carte mentale générale](#) (ici en archive [zip](#)) au sujet de la qualification des données suivant la

norme ISO 19157.

J-M. Arzac a initié l'élaboration de la nomenclature des usages génériques de l'information géographique (dont une ébauche était disponible dans la [note méthodologique du cas « Adresse »](#)). Cette nomenclature, qui doit être extensible, permettra de placer les priorités et pondération des critères et mesures de qualité en face des usages.

Pour le cas d'usage « Qualification des données d'adresses locales », S. Levêque présente : le [tableau synopsis](#) (en colonnes les critères et mesures de qualité, en ligne des éléments d'évaluation avec des points de vigilance tels l'évolution du modèle BAL, la dépendance à des données externes, etc.), l'[arbre de décision](#), et la [note d'accompagnement](#) précisant : le contexte, les définitions, les usages recensés, des commentaires sur le tableau synopsis, et la description des besoins pour aller plus loin.

Pour le cas d'usage « Qualification des données relatives aux équipements collectifs publics », S. Rolle présente le [tableau synopsis de priorisation et pondération des critères et mesures qualité](#) et l'[arbre de décision correspondant](#). S. Rolle suggère d'ajouter la méthode de calcul pour chaque critère, et A. Gallais suggère d'ajouter le cas d'utilisation « Accessibilité aux personnes à mobilité réduite ».

Les deux démarches constituent actuellement deux travaux de recherche offrant deux rendus différents constituant les livrables d'une démarche exploratoire. La finalité commune émergente consiste à déterminer comment on évalue la qualité des données en fonction de l'usage afin d'approcher une évaluation de la qualité externe des données.

#### Décision / Actions :

- *Rapprocher / harmoniser les livrables BAL et Equipco (S. Rolle, M. Rajerison, S. Levêque)*
- *Abonder la nomenclature des usages des usages génériques de l'information géographique*
- *Élaborer la note méthodologique générale avec une entrée par les usages de la part l'utilisateur, lui permettant d'identifier les critères et mesures utiles à observer pour cet (ces) usage(s) : (S. Levêque, N. Py)*

## **6. Restitution de la qualification des données géographiques**

Par S. Levêque. [La présentation est ici](#).

*Les diagrammes polaires offrent une excellente solution de restitution graphique de la qualité d'un jeu de données géographiques, 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 :*

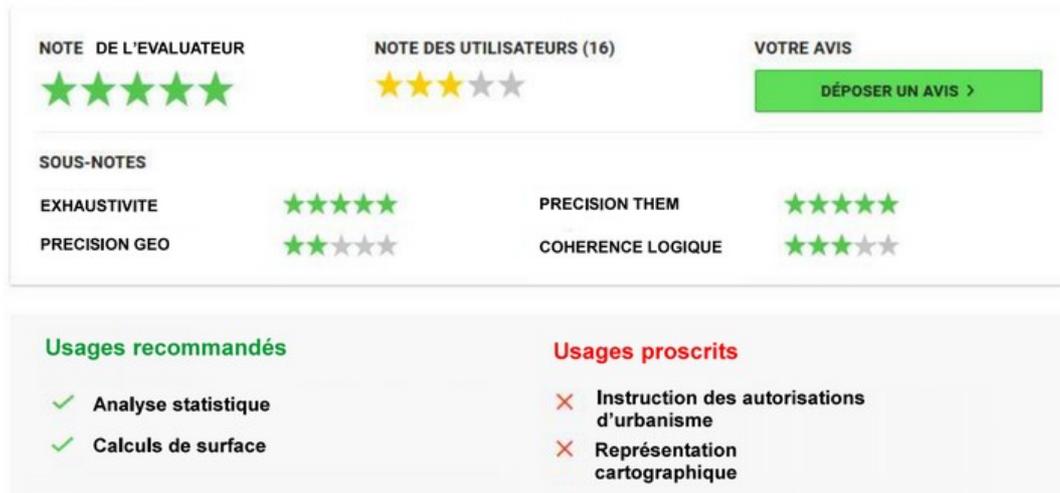
- *deux notes globales : la note du testeur (vert), et celle des utilisateurs (orange) ;*
- *une rubrique « déposer un avis » ;*
- *des sous-notes suivant des thèmes (ex : facilité d'obtention, facilité d'usage, « fraîcheur », etc.) ;*
- *une rubrique « points forts / points faibles », potentiellement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

Des exemples de restitution graphique de la qualification des données géographiques ont

été fournis au §2 (SI Milieu Marin) et au §3 (Qualification des réseaux).

En s'inspirant du site « Les numériques » et en prenant l'exemple d'un document d'urbanisme, S. Levêque a réalisé la maquette graphique de restitution de la qualification des données géographiques, en introduisant une note de l'évaluateur, une note de l'utilisateur (avec possibilité de déposer un avis !), une notation par critères qualité, et une qualification suivant l'usage en distinguant clairement ceux recommandés et proscrits.

### RÉSUMÉ : PLU de Numérac



Par ce principe, différents lots de données analogues (par exemple différents documents d'urbanisme) peuvent également être comparés entre eux suivant leurs critères de qualité (qualité interne) mais également suivant des usages (qualité externe), cf. [présentation diapo 3](#). Ce sujet confirme le besoin d'une nomenclature des principaux usages de l'information géographique.

#### Décision / Actions :

- Développer un outil simple (macro excel ou autre...) de saisie des évaluations et de génération du résultat graphique conforme à cette maquette. (S. Levêque, M. Rajerison)

#### Remarque générale :

A ce stade, la production effective ou en cours du GT CNIG QuaDoGéo comprend :

- Les fiches méthodologiques du Cerema ;
- Le registre français de mesures de qualité des données géographiques ;
- 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, et la méthodologie correspondante ;
- Une nomenclature des usages génériques de l'information géographique ;
- Des maquettes et outil de restitution de la qualité des données ;
- Une évaluation du GUF, outil de retour utilisateur quant à la qualité des données.