



**MANDAT POUR UN GROUPE DE TRAVAIL  
« QUALIFICATION DES DONNÉES GÉOGRAPHIQUES »  
DE LA COMMISSION RÈGLES DE MISE EN OEUVRE**

### **PRÉAMBULE**

Le Cerema et ses partenaires travaillent depuis plusieurs années sur la question de la qualification des données géographiques et de ses enjeux. Au niveau méthodologique, un décryptage des normes ISO, piloté par le Cerema, a donné lieu début janvier 2018 à la publication d'une série de fiches<sup>1</sup>. Pour le CRIGE-PACA, le contrôle de la qualité des données est un élément essentiel de sa plateforme, et il a organisé en juin 2017 à Aix-en-Provence une journée technique<sup>2</sup> permettant d'identifier les enjeux et les principaux acteurs.

Afin de poursuivre cette dynamique, le CRIGE-PACA et le Cerema ont organisé un séminaire, du 6 au 8 février 2018, à Lyon, sur le thème : « Quels outils et organisations pour qualifier les données géographiques ? ». Ce séminaire était à vocation opérationnelle, sur la mise en pratique de la qualification des données géographiques. En regroupant une quinzaine de structures, publiques comme privées, qui ont pu travailler sur la qualité des données, le séminaire a permis de constituer un groupe pour à la fois favoriser et mettre en cohérence les initiatives existantes. L'objectif initial était celui de la construction d'un plan d'actions.

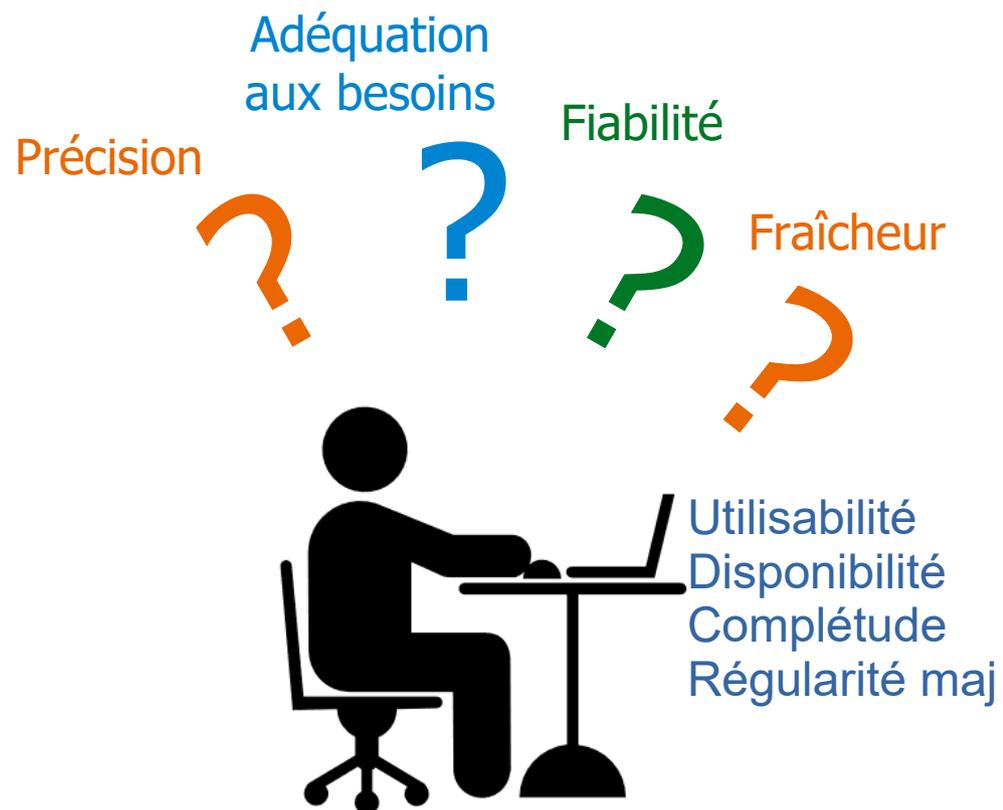
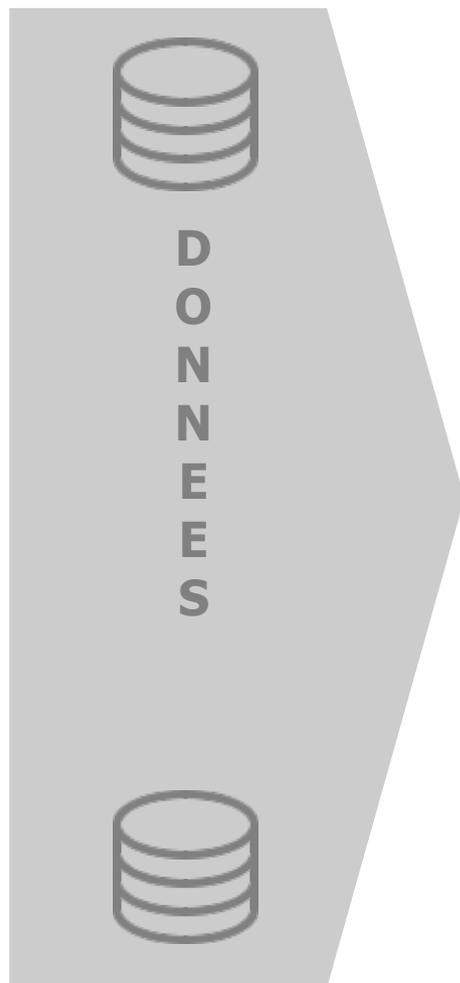
A l'occasion de ce séminaire est née l'idée d'un groupe de travail pour amplifier la dynamique amorcée, et généraliser l'appropriation des données géographiques, particulièrement les données ouvertes.



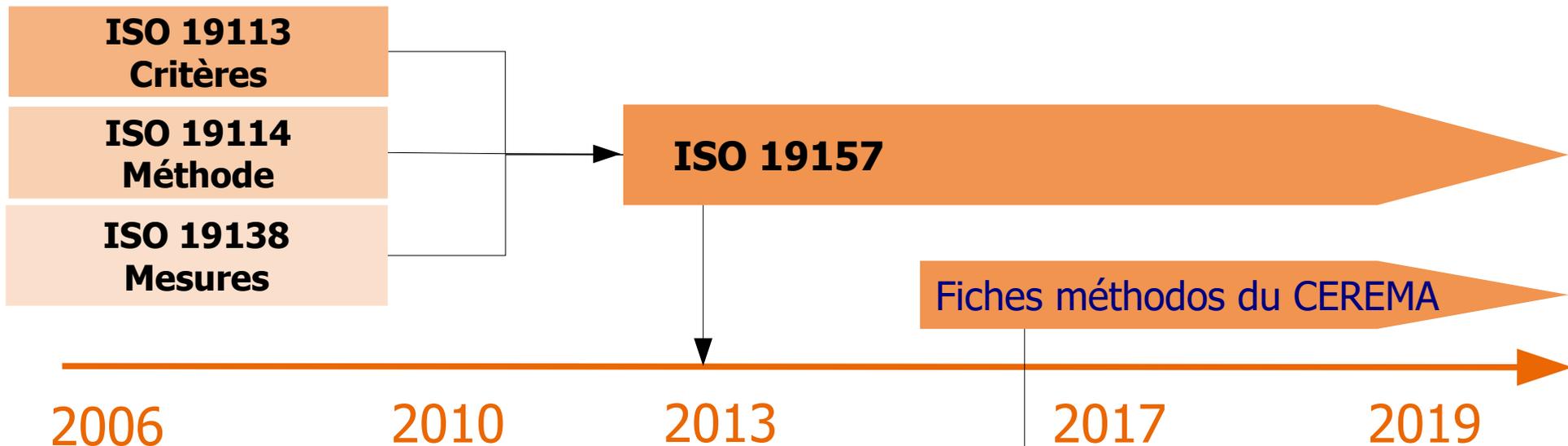
# Mandat 2019-20 du GT QuaDoGéo

- Enjeu / mission : « *aider à la réutilisation des données ouvertes grâce à une meilleure information sur leur qualité* »
- Axes de travail :
  - *accompagnement des diffuseurs et des utilisateurs de données*
  - *prise en compte du retour utilisateur (crowdsourcing, évaluation...)*
  - *exemple d'action concrète : synopsis pratique pour la qualification*
- Animation et secrétariat technique : Cerema (A. Gallais), avec le soutien du CGDD/SRI et de la Mission CNIG
- Composition du GT : collectivités, prestataires de services, IGN production, IGN cellule certification, OpenData France, IFREMER, Cerema, etc.
- Démarrage : mars 2019
- 4 réunions par an (=> 7 réunions réalisées + ateliers en sous-groupes)

# Enjeux de la qualification des données



# Bref rappel historique...



2006

2010

2013

2017

2019

Expertise de G. Troipoux

Bienvenue dans ce guide interactif sur la "qualité des données géographiques".



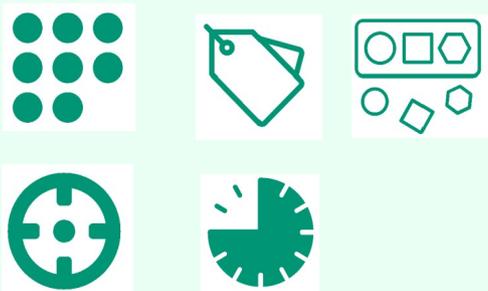
Commencer la lecture du guide par le commencement ! (Introduction, Contexte, Objet...)



# (rappel) Fiches méthodologiques

## Critères

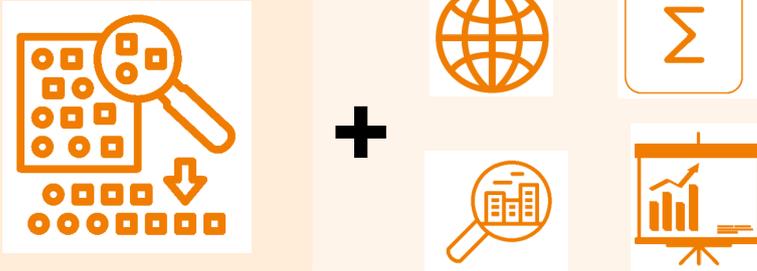
19157



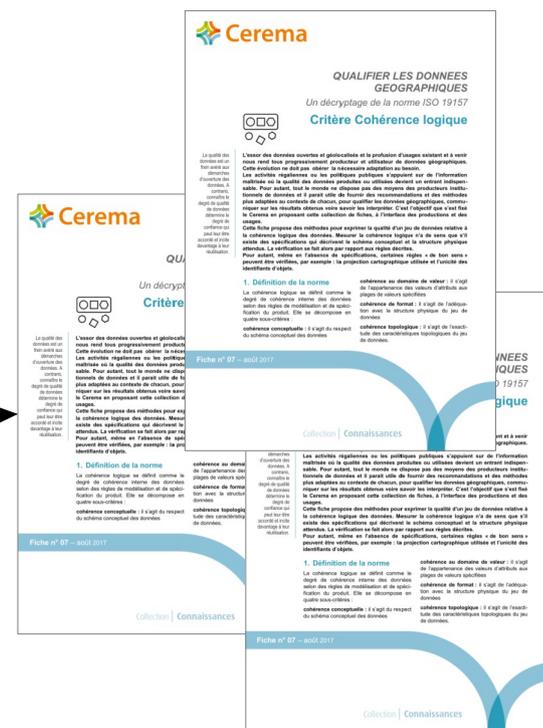
Exhaustivité  
Précision thématique  
Cohérence logique  
Précision de position  
Cohérence temporelle

## Méthodes

19157



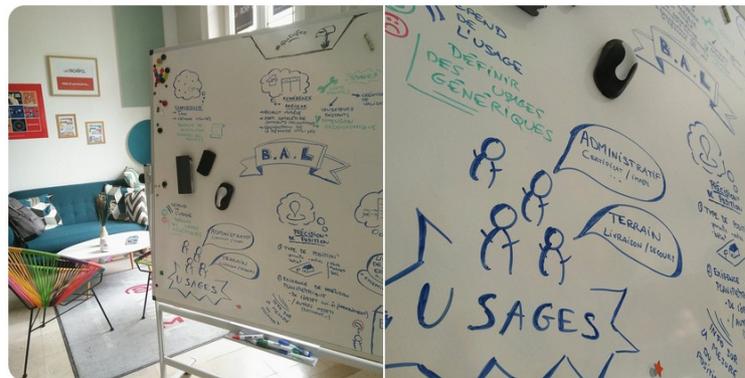
Echantillonnage  
Généralités sur la qualif.  
Contexte de contrôle  
Éléments statistiques  
Méthodes de représentation



# Bilan du groupe de travail 2019 - 2020

# Retours d'expérience

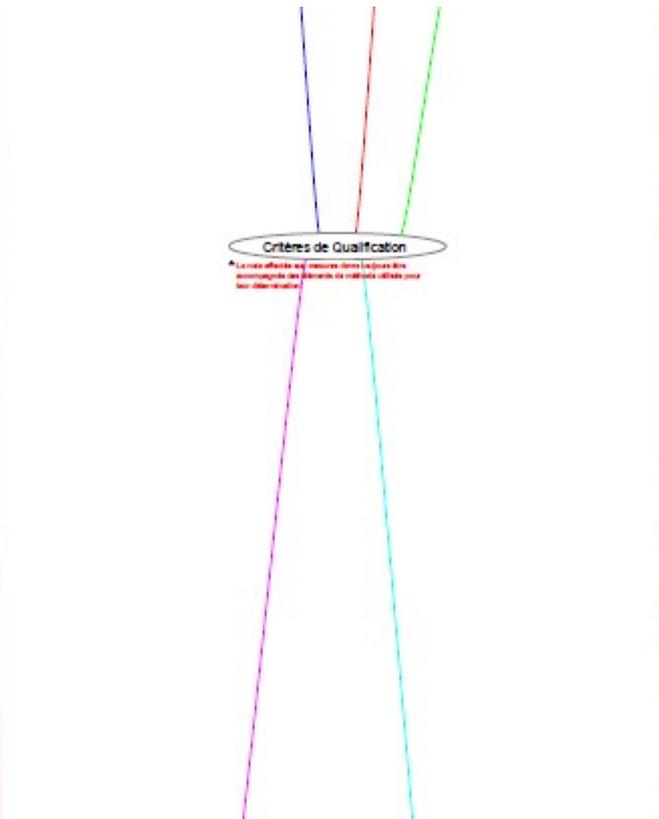
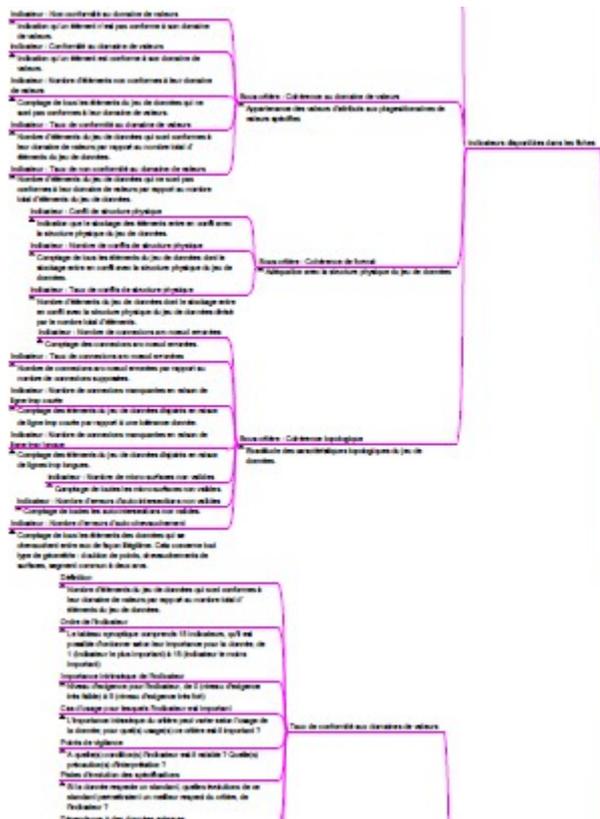
- Nombreux retours d'expérience (producteurs / utilisateurs)
  - Contrôles qualité chez Axes Conseil
  - Les contrôles qualité de l'IGN
  - Qualification de la donnée du Système d'Information sur le Milieu Marin (SIMM)
  - Qualification de l'OCSGE à la métropole de Nice Côte d'Azur
  - Qualification de la précision de la BAN par analyse géostatistique par IDéo Bourgogne Franche-comté
  - Présentation du Géospatial User Feedback (GUF)
  - etc..
- Communication
  - Site CNIG et Infos CNIG
  - Cerema, Georezo, Twitter
  - Participation (N.Py, J.Marc au Q-KEN)



04:05 - 8 avr. 2019

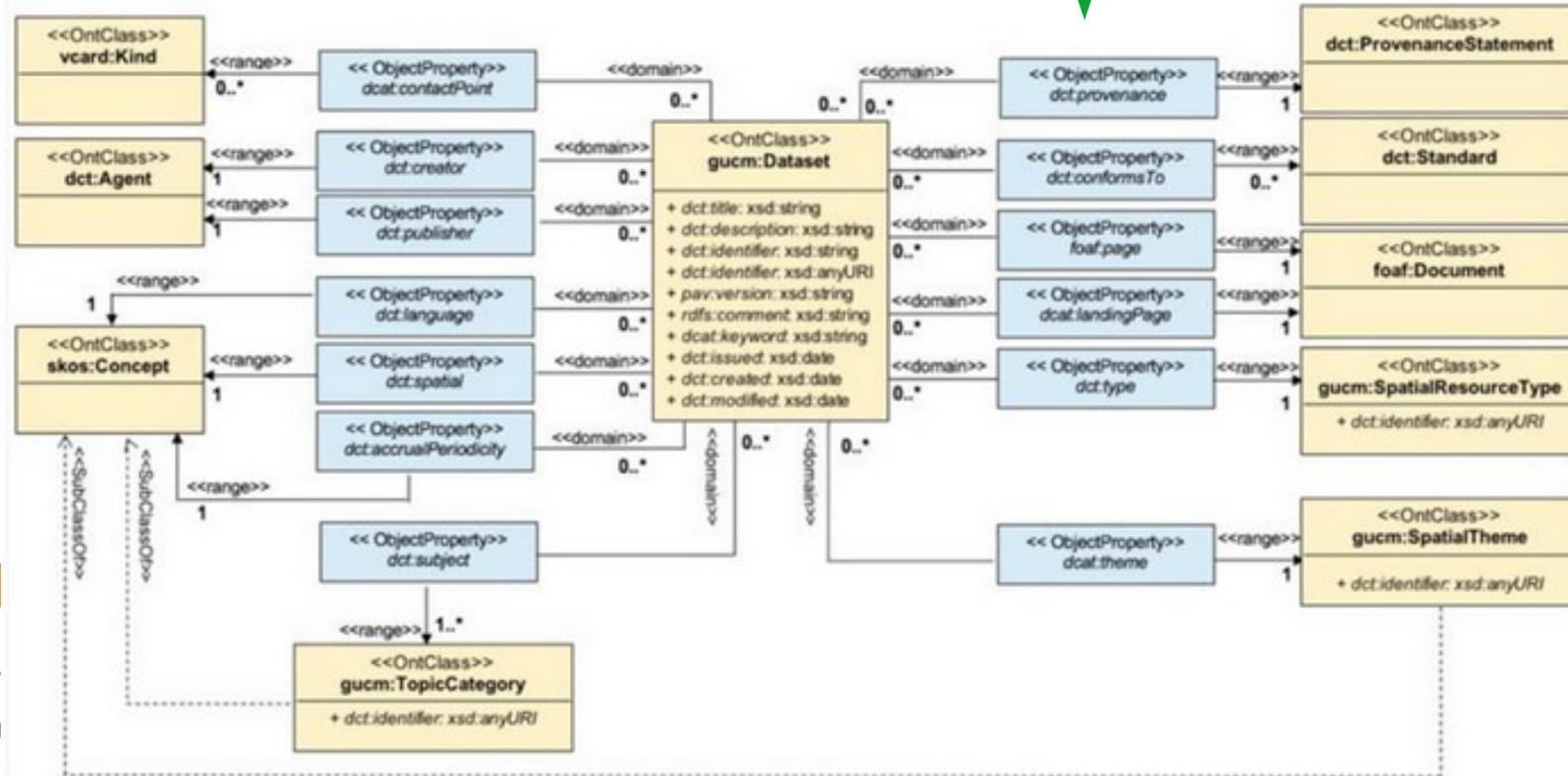
# Décryptage ISO19157

- Principalement N. Py + cellule normalisation
  - Carte mentale : qualification des données suivant ISO1957
  - Norme ISO 19157 au format XML sur Github



# Veille méthodologique

- Veille sur les actions de recherches internationales
  - Geographic Information Metadata - Outlook from the International Standardization Perspective
  - Reducing Consumer Uncertainty: Towards an Ontology for Geospatial User-Centric Metadata (ontologie pour les métadonnées géospatiales centrées utilisateur (GUCM))
- Veille sur le principe du retour utilisateur
  - Geospatial User Feedback (GUF)
- Veille sur open data et data.gouv
  - (à venir.. !) quelles métadonnées qualité sur les portails open data ?
  - *schema.data.gouv*



# Registre français des MD de qualité de l'IG

- Objectif: améliorer la saisie des critères qualité dans les métadonnées ISO
- Il peut constituer une « *extension pour les éléments de qualité* » au guide CNIG des métadonnées v2.0
- Implémenté sur le « Syst. de publication de registres » du Géocatalogue
  - Chaque mesure qualité est désormais référencée par un URL :

## 26 mesures

21 mesures ISO19157 retenues dans les fiches méthodo du Cerema

2 mesures hors ISO19157 mais préconisées dans les fiches Cerema :  
- coeff. Kappa (just. de classement)  
- classe de précision de l'arrêté 2003

3 éléments de qualité de ISO19115 :  
- généalogie  
- résolution spatiale  
- cohérence topologique



Métadonnées relatives à la  
qualité des données géographiques

## REGISTRE FRANÇAIS

**Version Projet**  
**Registre CNIG**  
**octobre 2019**

Système de publication de registres

Explorer À propos Avancé Rechercher Soumettre

https://data.geocatalogue.fr/nol/\_mesuresQuaDoGeo

Registre: mesuresQuaDoGeo  
URI: https://data.geocatalogue.fr/nol/mesuresQuaDoGeo  
Mesures liées à la Qualité de Données Géographiques

Métadonnées principales  
Toutes les propriétés  
Télécharger  
Envoyer commentaire

Contenu

Voir 20 lignes

Filter les entrées:

Nom	Notation	Description	Types	Statut
Classe de précision au sens de l'arrêté du 16 septembre 2003	10	L'arrêté du 16 septembre 2003 porte sur les classes de précis...	dQM measure , Concept	experimental
Coefficient Kappa	16	Matrice indiquant le nombre d'éléments de la classe (i) class...	dQM measure , Concept	experimental
Conflit de structure physique	4	Indication que le stockage des éléments entre en conflit avec...	Concept , dQM measure	experimental
Date de création	24	Date de création de la ressource.	dQM measure , Concept	experimental
Date de publication	26	Date de publication de la ressource.	dQM measure , Concept	experimental
Date de révision	25	Date de dernière révision de la ressource.	dQM measure , Concept	experimental
Erreur circulaire avec un seuil de signification de 95 %	12	Rayon décrivant un cercle, dans lequel l'emplacement vérita...	dQM measure , Concept	experimental
Erreur horizontale relative	14	Evaluation des erreurs aléatoires d'une entité dans la positi...	dQM measure , Concept	experimental
Erreur verticale relative	13	Evaluation des erreurs aléatoires d'une entité du relief à un...	dQM measure , Concept	experimental
Etendue temporelle	23	L'étendue temporelle définit la période de temps ouverte par...	dQM measure , Concept	experimental
Exactitude de la mesure temporelle avec un seuil de signification de 95 %	19	Demi-longueur de l'intervalle défini par une limite supérieur...	Concept , dQM measure	experimental
Généalogie	21	La généalogie fait état de l'historique du traitement et/ou d...	dQM measure , Concept	experimental
Incertitude de la valeur d'attribut avec un seuil de signification de 95 %	18	Demi-longueur de l'intervalle défini par une limite supérieur...	Concept , dQM measure	experimental
Matrice relative de classement erroné (RMCM)	15	Matrice indiquant le nombre d'éléments de la classe (i) class...	Concept , dQM measure	experimental
Nombre de micro-surfaces non valides	7	Comptage de tous les éléments du jeu de données qui constitue...	Concept , dQM measure	experimental
Nombre d'erreurs de chevauchement	9	Comptage de tous les éléments du jeu de données qui se chevau...	Concept , dQM measure	experimental

# Processus de qualification des données

- Deux cas étudiés : les bases adresses locales, une base des ERP
- Des ateliers remue-méninges



- ➔ Une méthode pratique
- ➔ Un arbre de décision
- ➔ Un tableau synopsis

GT CNIG - Qualification des données géographiques

## Méthode pratique pour qualifier des données version de travail, 5 juin 2020

### Préambule

Le présent document constitue une méthode pratique pour qualifier les données géographiques. Il vise à proposer des éléments de mise en œuvre des [fiches méthodologiques](#) produites par le Cerema, et est réalisé dans le cadre du [groupe de travail du CNIG](#).

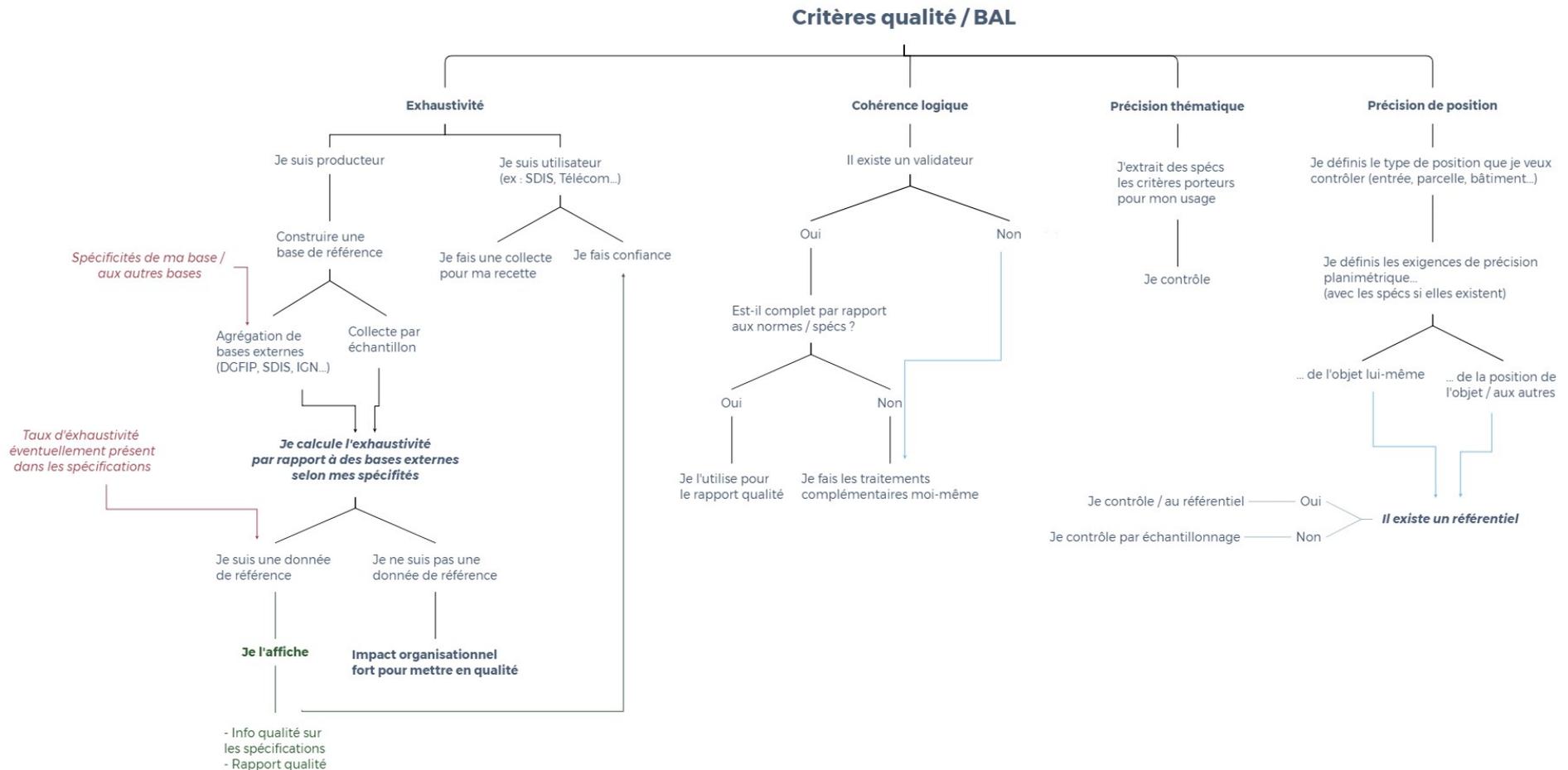
Les différentes étapes de la méthode sont illustrées par deux cas pratiques : la base adresse locale et les équipements collectifs.

### Sommaire

1. Définir le périmètre et les objectifs.....	2
1.1 Que qualifie-t-on ?.....	2
1.2 Pourquoi ?.....	3
2. Cerner les usages.....	4
2.1 Thèmes.....	4
2.2 Usages.....	4
3. Hiérarchiser les critères qualité.....	6
4. Représenter les résultats.....	7
ANNEXES.....	8

# Processus de qualification des données

- Synopsis pratique, sous la forme : 1/ d'un arbre de décision



# Processus de qualification des données

- Synopsis pratique

- 2/ tableau de synthèse de qualification des bases adresses locales
- pondération et priorisation des critères de qualification

BAL	Exhaustivité	Cohérence logique						Précision thématique			Précision de position		Qualité temporelle		
	Taux d'exhaustivité	Taux de conformité aux domaines de valeurs	Conflit de structure physique	Taux de connexions erronées	Nb de micro-surfaces non valides	Nb d'auto-intersections non valides	Nb de chevauchements	Matrice relative de classement erroné	Taux de valeurs d'attribut correctes	Incertitude avec seuil de signification de 95 %	Valeur moyenne des incertitudes (écarts)	Classe de précision (arrêté 16/09/2003)	Exactitude avec seuil de signification de 95 %	Taux de conformité chronologique	Taux de conformité aux domaines de valeurs
Importance intrinsèque du critère	4	3 pour les attributs « position » (entrée, délivrance postale...), « suffixe », « numéro », et présence du point dans l'emprise géographique de la commune	3	NC	NC	NC	NC	3 – Attributs « position » et « suffixe » (le remplissage erroné n'est pas bloquant, car il dégrade juste la précision de position.)	4 ou 3 – Notamment 3 pour NOM COMMUNE / NOM VOIE	NC	3 (avec des classes de précisions potentiellement définies par les specs locales)	NC	1 – DATE DE DERNIERE MISE A JOUR	NC	1
Cas d'usage pour lesquels le critère est important (Administratif, Défense / secours, vie quotidienne, technique / gestionnaire)	SECOURS GESTIONNAIRE	TECHNIQUE (rattachement à une parcelle par relation spatiale)	TOUS	NC	NC	NC	NC	Usages pour lesquels la précision de position est importante	TOUS	NC	Important dès lors que pour une même adresse, il y a plusieurs « positions ».	NC	RAS	NC	RAS
Ordre de chaque critère	3	2	1	NC	NC	NC	NC	RAS		NC		NC	RAS	NC	RAS
Points de vigilance	Respecter le standard : filaire de voies avec bâti exhaustif (les voies sans adresse doivent être présentes)	RAS	RAS	NC	NC	NC	NC	RAS	RAS	NC	Faire une acquisition de contrôle correspondant à la précision nécessaire de l'attribut « position » (ex : entrée bâtiment : précision métrique)	NC	RAS	NC	RAS
Pistes d'évolution des spécifications	Lieux-dits à préciser (présence obligatoire d'un numéro ?)	Prise en compte plus large des pratiques locales sur la position (ex : position = « secours ») Rendre obligatoire la présence des coordonnées dans un des deux systèmes. Imposer l'égalité des coordonnées si elles sont présentes dans les deux systèmes.	RAS	NC	NC	NC	NC	RAS	Préciser la manière d'écrire les noms de rues (pas d'abréviation, D' ou DE...) => utiliser l'une des données externes ?	NC	INTEGRER AU MODELE DES ATTRIBUTS RELATIF A LA QUALITE (EX : INCERTITUDE SUR LA POSITION, COMME DANS LA BD TOPO)  PREVOIR D'INTEGRER DES SEUILS DE QUALITE DANS LES SPECS	NC	COHERENCE DES DATES DE MAJ AVEC LES CREATIONS DE RUES... COHERENCE INTERNE DE LA MISE A JOUR DES ADRESSES D'UNE MEME RUE  Intégrer la date de dernière vérification (« confirmation »)	NC	RAS
Dépendance à des données externes	Filaire de voie : Base de comparaison possible : adresses fiscales	Registre de valeur du standard pour « position » et « suffixe »	Schéma du validateur de la BAL	NC	NC	NC	NC	TYPE FANTOIRE REGISTRE DES VOIES COG ADMIN EXPRESS			RAS	NC	RAS	NC	RAS
Besoin d'outils complémentaires	RAS	Besoin d'amélioration des contrôles du validateur sur la position géographique.	RAS	NC	NC	NC	NC	RAS	Registre des voies par commune (en API ?)	NC	RAS	NC	RAS	NC	RAS

# Restitution sur la qualité des données

- Prototype d'un outil de restitution graphique de la qualification

– Inventaire de solutions graphiques :



– Création de la maquette :

## RÉSUMÉ : PLU de Numérac

<b>NOTE DE L'EVALUATEUR</b> ★★★★★	<b>NOTE DES UTILISATEURS (16)</b> ★★★★☆	<b>VOTRE AVIS</b> <a href="#">DÉPOSER UN AVIS &gt;</a>
--------------------------------------	--------------------------------------------	-----------------------------------------------------------

**SOUS-NOTES**

<b>EXHAUSTIVITE</b> ★★★★★	<b>PRECISION THEM</b> ★★★★★
<b>PRECISION GEO</b> ★★★☆☆	<b>COHERENCE LOGIQUE</b> ★★★☆☆

<b>Usages recommandés</b>	<b>Usages proscrits</b>
✓ Analyse statistique	✗ Instruction des autorisations d'urbanisme
✓ Calculs de surface	✗ Représentation cartographique

# Restitution sur la qualité des données

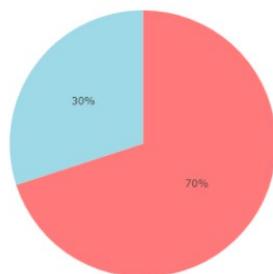
- Prototype d'un outil de restitution graphique de la qualification
  - développement en cours d'un outil / POC : *quadorender.py*
    - mesures de qualité (indicateurs ISO 19157)
    - notes de qualité calculées depuis les métriques
    - notes de qualité selon les usages

## positionAccuracy

ClassePrecisionArrete2003 : Classe 1

valMoyIncertitude : 0.3

valMoyIncertitude



■ Non conforme

■ Conforme

## Spatial statistics and reports

exhaustivity : ★★★★★★★★★★

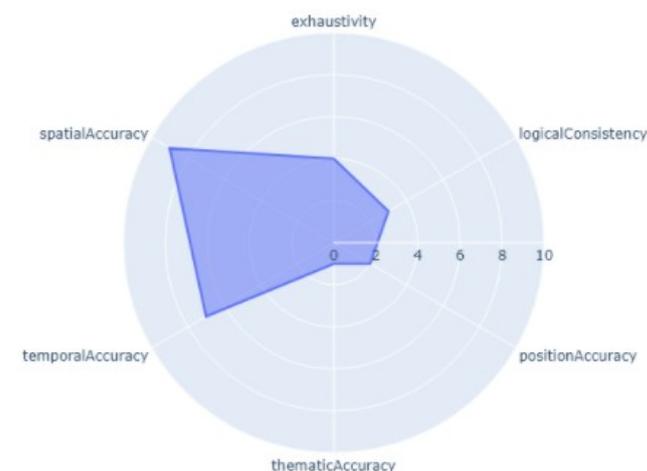
logicalConsistency : ★★★★★★★★★★

positionAccuracy : ★★★★★★★★★★

thematicAccuracy : ★★★★★★★★★★

temporalAccuracy : ★★★★★★★★★★

spatialAccuracy : ★★★★★★★★★★



# Tableau de bord des actions

Retours d'expérience		Actions en continu
Communication		Accentuer la communication
Veille méthodologique		
Registre français des MD de qualité de données géographiques		Projet de registre à valider / publier
Processus de qualification des données		Fiche méthode à publier
Restitution graphique sur la qualité des données		Poursuivre le développement

# Conclusions et pistes de travail

- Sujet « Qualité » toujours aussi distrayant... 😊
  - on essaye de l'aborder par des retours d'expériences et des outils ou solutions pratiques
  - Le stock de RETEX s'épuise un peu => inciter à de nouveaux retours d'expérience
- Registre français des métadonnées de qualité des données géographiques
  - Le valider / diffuser
  - Revoir l'hébergement du registre ?
- Rendre les MD sur la qualité externe moins facultatives dans les géocatalogues français
- Evaluer les géostandards au regard de la thématique qualité
  - Rattachement à la Commission Données ?
- Accentuer la veille sur les outils de retour des utilisateurs
  - utilisation du GUF ?
- Poursuivre le développement de l'outil de restitution graphique de la qualification, en ciblant la réalisation de la maquette graphique.



**Cerema**

Centre d'études et d'expertise sur les risques,  
l'environnement, la mobilité et l'aménagement

Direction territoriale Ouest

**Merci de votre attention**

Arnauld GALLAIS

+33 (0)2 40 12 84 76

[arnauld.gallais@cerema.fr](mailto:arnauld.gallais@cerema.fr)