



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 fiches¹. 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 technique² 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 principal était celui de la construction d'un plan d'actions.

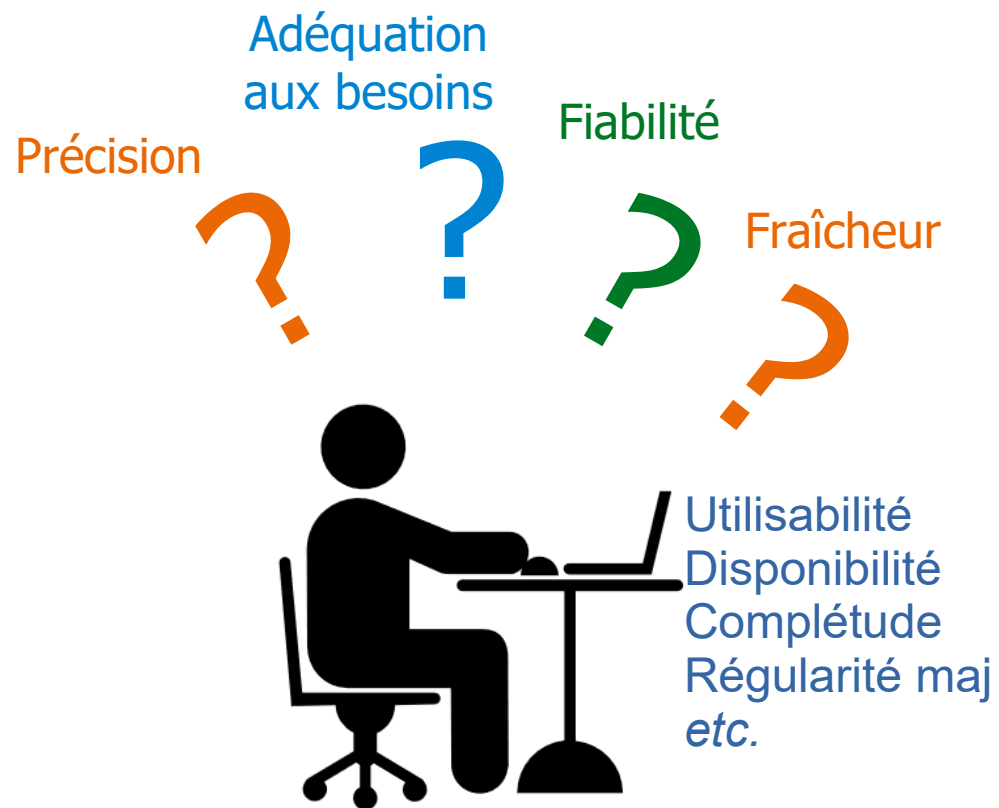
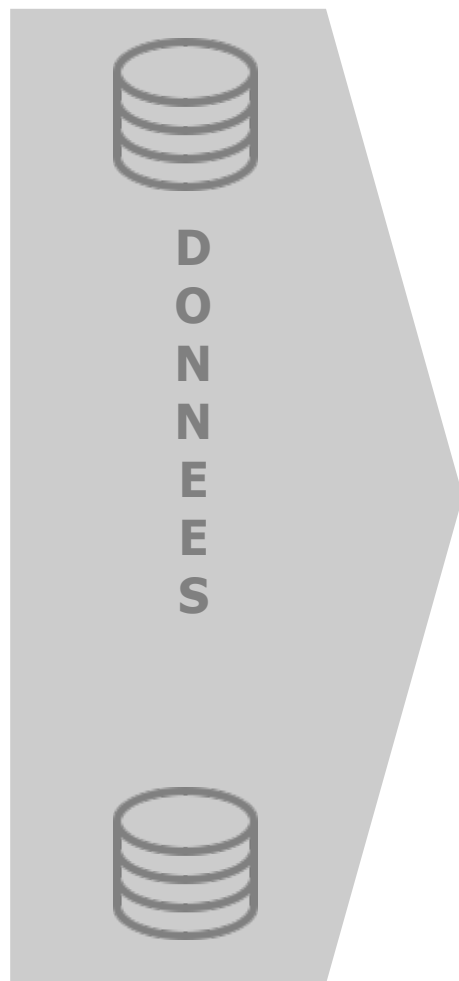
A l'occasion de ce séminaire est née l'idée d'un groupe de travail pour amplifier la dynamique amorcée, et généraliser l'appropriation des données géographiques, particulièrement les données ouvertes.



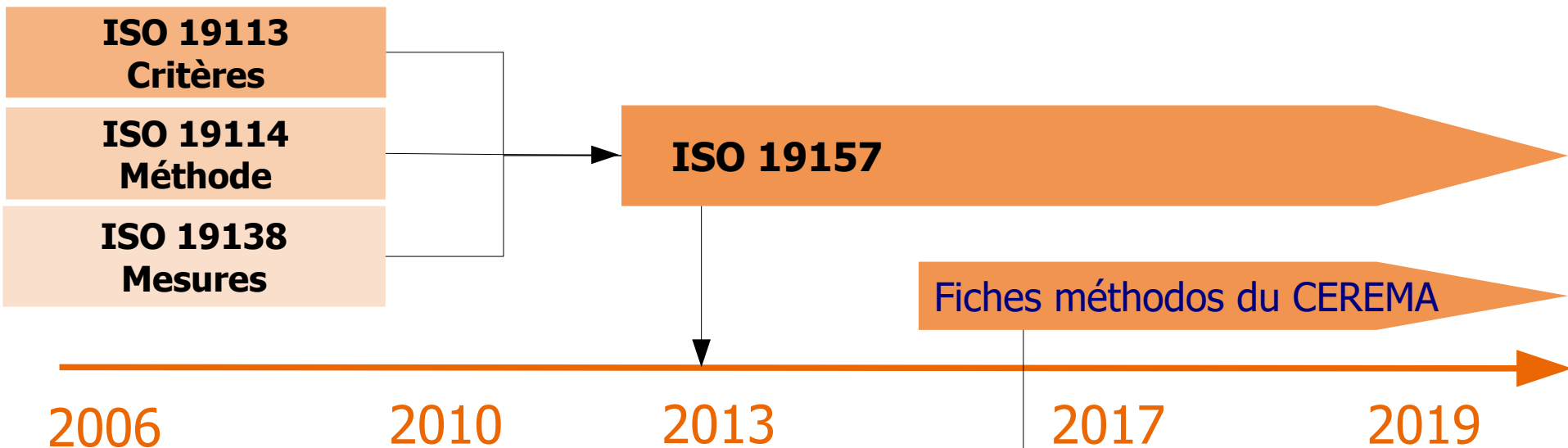
Mandat 2019-20 du GT QuaDoGéo

- 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...)*
 - *actions concrètes : synopsis pratique pour la qualification, registre des mesures de qualité*
- 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 / an + ateliers en sous-groupes

Enjeux de la qualification des données



Bref rappel historique...



2006

2010

2013

2017

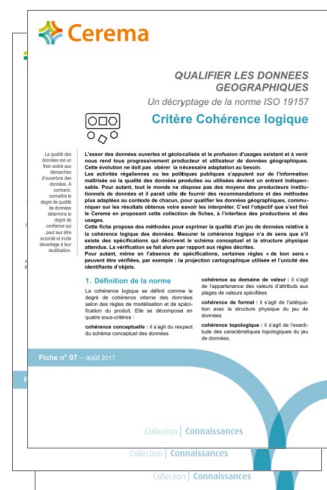
2019

Expertise de G. Troipoux

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



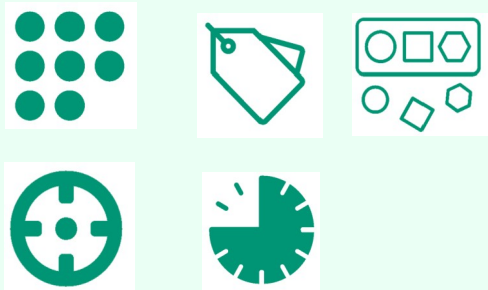
Commencer la lecture du guide par le commencement ! (Introduction, Contexte, Objet...)



Fiches méthodologiques

Critères

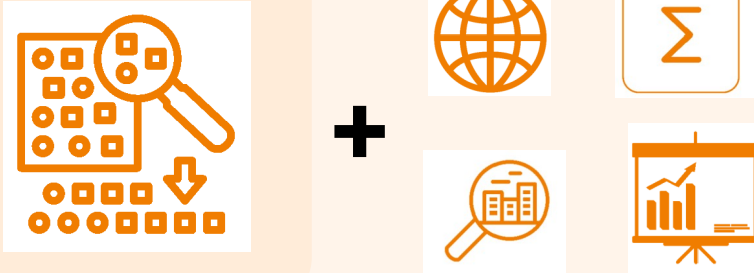
19157



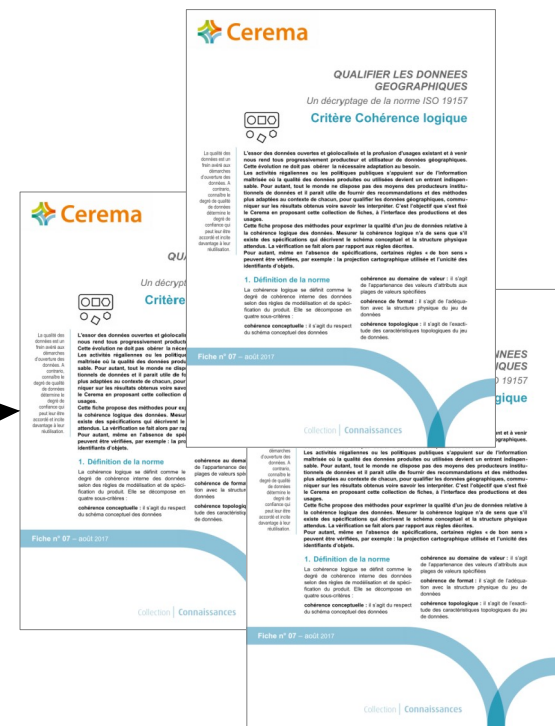
Exhaustivité
Précision thématique
Cohérence logique
Précision de position
Cohérence temporelle

Méthodes

19157



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

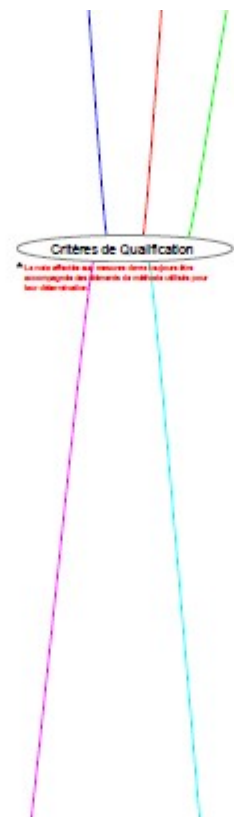
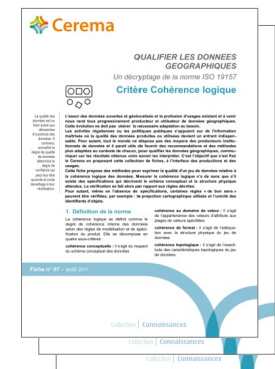


Actions du groupe de travail

2019 - 2020

Décryptage ISO19157

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



Retours d'expérience

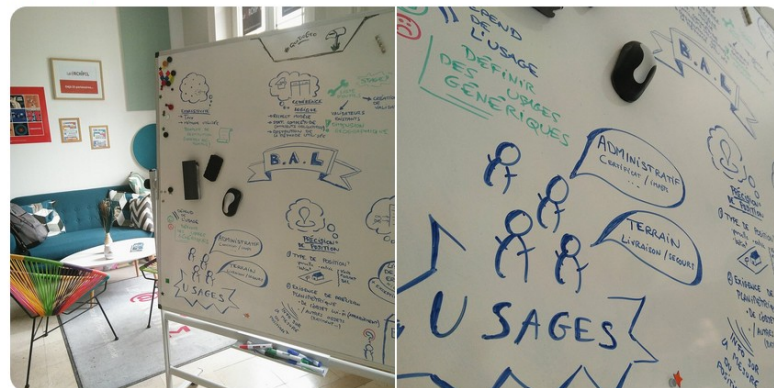
- Nombreux retours d'expérience (producteurs / utilisateurs)
 - Contrôles qualité chez Axes Conseil
 - Les contrôles qualité de l'IGN
 - Qualification de la donnée du Système d'Information sur le Milieu Marin (SIMM)
 - Qualification de l'OCSGE à la métropole de Nice Côte d'Azur
 - Qualification de la 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
 - Participations (N.Py, J.Marc au Q-KEN)



Stéphane Lévêque
@tendanceville

Suivre

Matinée brainstorming lyonnais sur la qualité des données géographiques, dans le cadre du travail du @CNIG_France sur @quadogeo | Merci @Lab_Archipel pour l'accueil ! | On envoie vite un CR de nos réflexions à tout le groupe ;-)

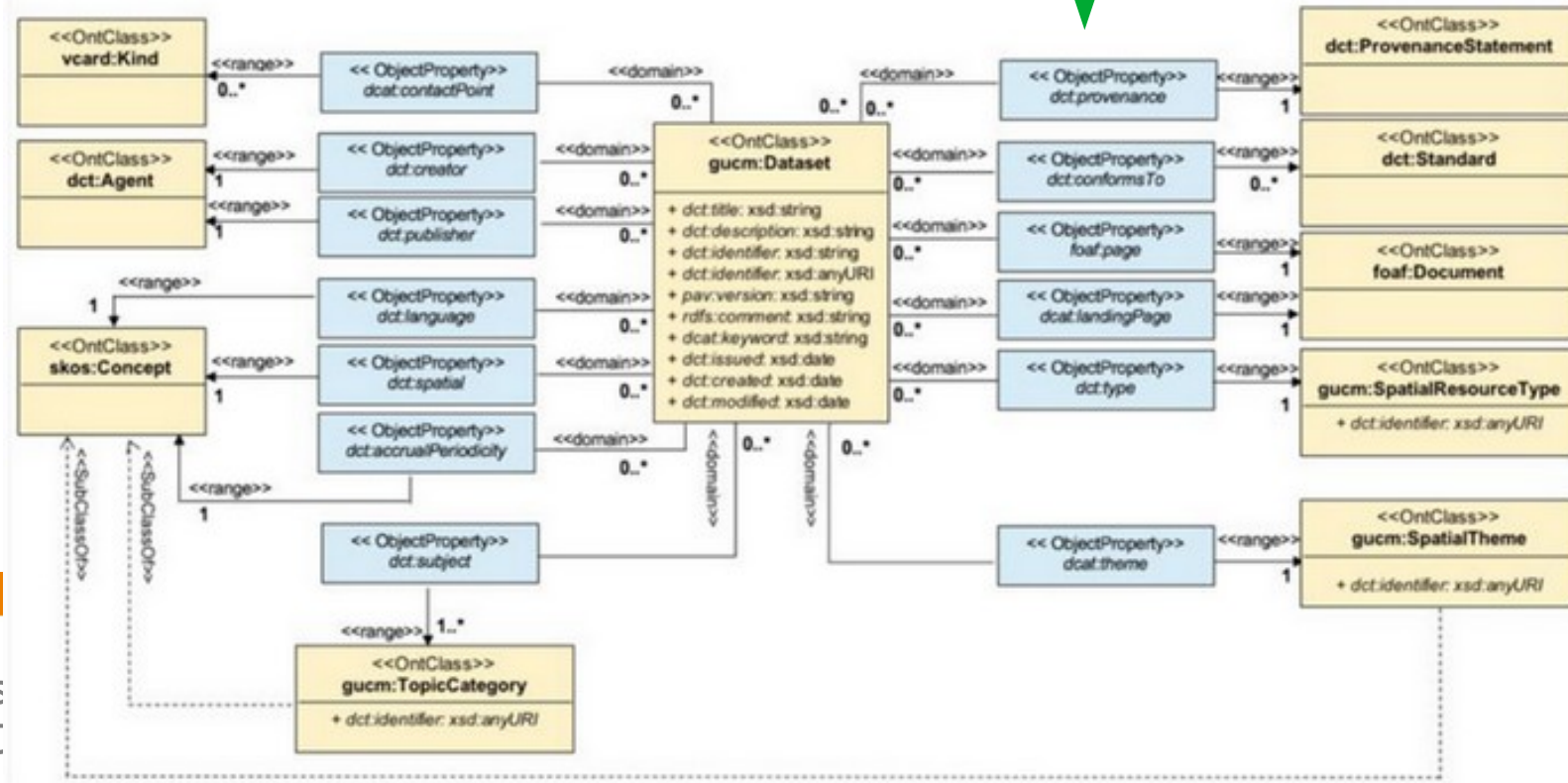


04:05 - 8 avr. 2019



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
 - quelles métadonnées sur la qualité dans les portails open data ?



Registre des mesures de qualité de l'IG

- Objectif: améliorer la saisie des critères qualité dans les métadonnées ISO
- Conçu pour constituer une « *extension pour les éléments de qualité* » au guide CNIG des métadonnées v2.0



27 mesures

22 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

Registre des mesures de qualité de l'IG

- Implémenté sur le « Syst. de publication de registres » du Géocatalogue
 - Chaque mesure qualité est référencée par un URL :

Système de publication de registres géo catalogue

[Explorer](#) [À propos](#) [Avancé ▾](#)

https://data.geocatalogue.fr/ncl/_mesuresQuaDoGeo experimental

Registre: mesuresQuaDoGeo Métadonnées principales

URI: <https://data.geocatalogue.fr/ncl/mesuresQuaDoGeo> Toutes les propriétés

Mesures liées à la Qualité de Données Géographiques Télécharger

Envoyer commentaire

Contenu

Voir lignes Filtrer les entrées:

Nom	#	Description	Types	Statut
Taux d'exhaustivité	1	La formule de calcul du taux d'exhaustivité figure page 2/6 d...	Concept , dQM measure	experimental
Non conformité aux règles du schéma conceptuel	2	Si le schéma conceptuel décrit implicitement ou explicitement...	Concept , dQM measure	experimental
Taux de conformité au domaine de valeurs	3	Résultat de la moyenne des taux de conformité des attributs p...	dQM measure , Concept	experimental
Conflit de structure physique	4	Indication que le stockage des éléments entre en conflit avec...	Concept , dQM measure	experimental
Taux d'erreur de formatage	5	Certaines valeurs doivent respecter des formats bien spécifiqu...	dQM measure , Concept	experimental
Taux de connexions erronées	6	Il existe une connexion nœud-tronçon là où différents tronçon...	Concept , dQM measure	experimental

Registre des mesures de qualité de l'IG

Entrée: Taux d'exhaustivité

URI: <https://data.geocatalogue.fr/ncl/mesuresQuaDoGeo/txEx>

definition	Nombre total d'éléments en excès ou manquant dans le jeu de données par rapport au nombre total d'objets du jeu de données.
element name	Exhaustivité / Taux d'exhaustivité
name	Taux d'exhaustivité
parameter	-
source reference	Fiche méthodologique Critère d'exhaustivité Non référencé dans ISO 19157
value structure	-
value type	Nombre réel, pourcentage
basic measure	Taux d'erreur
definition	Nombre total d'éléments en excès ou manquant dans le jeu de données par rapport au nombre total d'objets du jeu de données.
description	La formule de calcul du taux d'exhaustivité figure page 2/6 de la Fiche n°08 : Critère d'exhaustivité" de la série de fiches Cerema "Qualifier les données géographiques - Un décryptage de la norme ISO 19157" accessible en suivant ce lien : https://www.cerema.fr/fr/actualites/serie-fiches-cerema-qualifier-donnees-geographiques. "
notation	1
pref label	Taux d'exhaustivité

Processus de qualification des données

- Synopsis pratique

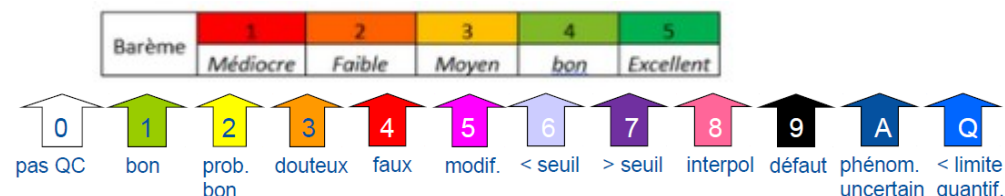
- 2/ tableau de synthèse de qualification des bases adresses locales
- pondération et priorisation des critères de qualification

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

NOTE DE L'EVALUATEUR ★★★★★	NOTE DES UTILISATEURS (16) ★★★★☆	VOTRE AVIS DÉPOSER UN AVIS >
--------------------------------------	--	--

SOUS-NOTES

EXHAUSTIVITE ★★★★★	PRECISION THEM ★★★★★
PRECISION GEO ★★★☆☆	COHERENCE LOGIQUE ★★★☆☆

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

Restitution sur la qualité des données

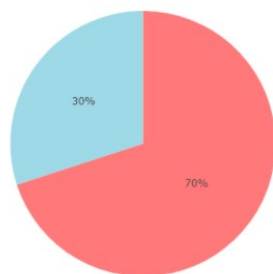
- Prototype d'un outil de restitution graphique de la qualification
 - développement en cours d'un outil / POC : *quadorender.py*
 - mesures de qualité (indicateurs ISO 19157)
 - notes de qualité calculées depuis les métriques
 - notes de qualité selon les usages

positionAccuracy

ClassePrecisionArrete2003 : Classe 1

valMoyIncertitude : 0.3

valMoyIncertitude



■ Non conforme

■ Conforme

Spatial statistics and reports

exhaustivity : ★★★★★★★★★★

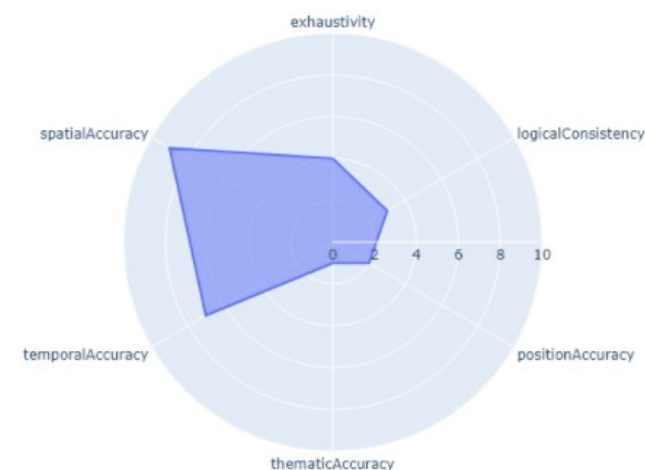
logicalConsistency : ★★★★★★★★★★

positionAccuracy : ★★★★★★★★★★

thematicAccuracy : ★★★★★★★★★★

temporalAccuracy : ★★★★★★★★★★

spatialAccuracy : ★★★★★★★★★★



Quels liens avec la Co Données ?

- Sujet « Qualité » toujours aussi... sexy...
 - on essaye de l'aborder par des retours d'expériences et des outils ou solutions pratiques
- Evaluer les géostandards au regard de la thématique qualité
 - vérifier la présence de spécifications de qualité dans tout nouveau ou nouvelle version d'un standard CNIG soumis à la validation de la Commission Données
 - faire inclure une sélection et une hiérarchisation des critères de qualité sous forme de propositions de seuils minimum dans les géostandards.
 - outre le critère « cohérence logique » (cf. validateur GPU, validata, etc.), ces seuils participeraient à l'évaluation de la conformité d'un lot de données par rapport à un standard
- Rendre les métadonnées sur la qualité externe « moins facultatives » dans les géocatalogues français
- Accentuer la veille sur les outils de retour des utilisateurs
- GT CNIG QuaDoGéo : centre de ressources méthodologique.



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