



Cerema

Démocratisons la qualité des données géographiques !



Nouvelles façons de penser

Gilles Troispoux – Cerema Territoires et villes
Yves Bonin – Cerema Méditerranée



Cerema Méditerranée

24 Mars 2017 - Decryptageo

Sommaire

- De quoi parle-t-on ?
- Une problématique déjà ancienne qui peine à mobiliser
- Un constat d'échec pour les utilisateurs
- Des réponses existent mais sont-elles adaptées?
- Nécessité d'avoir une approche plus pragmatique
- Une pensée pour les utilisateurs
- Et après?

De quoi parle-t-on ?

- Quels sont les besoins réels ?
- Pourquoi parle-t-on de qualité ?
- A-t-on besoin de données parfaites ?
- Ou simplement de connaître l'état des données ?
- Quelques éléments de réponse
- Les données parfaites n'existent pas ou très peu
- On cherche les données qui répondent à un besoin
- Nos ambitions sont parfois limitées
- L'analyse spatiale doit être maîtrisée
- Un indicateur doit être encadré
- Connaître l'incertitude sur les données est souvent indispensable

Une problématique déjà ancienne, mais qui peine à mobiliser

- **Des travaux menés par l'IGN**
 - Mise en place d'un projet qualité des données en 1991 à l'IGN
 - Travaux internationaux au niveau de l'ISO 2001 (ISO 19113, 19114, 19138)
- **Deux réseaux de recherche internationaux : ISSDQ, Accuracy**
- **Des réflexions menées par le CERTU à partir de 2011**
 - Administration des données, métadonnées
 - Didacticiel sous forme de CD Rom
 - Participation au réseau de recherche ISSDQ
 - Participation à un programme de recherche canadien
 - Participation au GT « Qualité des données » d'Inspire
- **Norme ISO 19157 en 2014**
- **Des réflexions menés par le réseau des CRIGE de l'AFIGEO**
 - Quelques outils pour la cohérence logique

Un constat d'échec pour les utilisateurs

- Un seul outil : la norme ISO 19157
 - Exclusivement réservée aux experts
 - Nécessite des ressources et de moyens importants
 - Nécessite que les données soient spécifiées initialement
- La qualité c'est chronophage et onéreux
- Spécificité de l'information géographique
- La crainte de diffuser de mauvais résultats
- Inspire
- l'Open Data
- Besoin d'outils simples et adaptés...

Des solutions existent ...



COVADIS



... mais sont-elles vraiment adaptées?

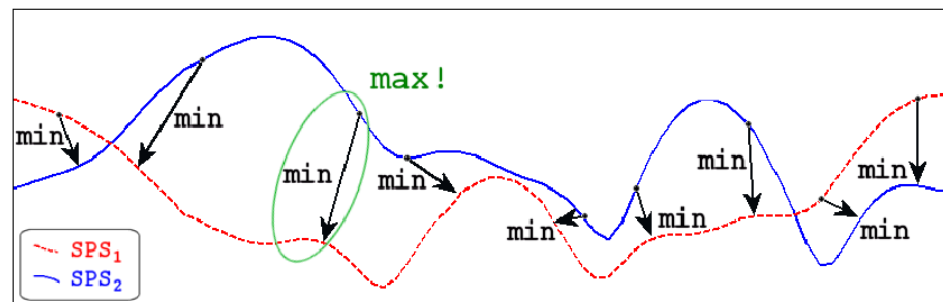
ISO 19157:2013(F)

Sommaire	Page
Avant-propos	iv
Introduction	v
1 Domaine d'application	1
2 Conformité	1
3 Références normatives	1
4 Termes et définitions	2
5 Abréviations	4
5.1 Abréviations	4
5.2 Abréviations relatives aux paquetages	5
6 Aperçu de la qualité de données	5
7 Composants de la qualité des données	6
7.1 Aperçu des composants	6
7.2 Unité de qualité des données	7
7.3 Éléments de qualité des données	8
7.4 Descripteurs des éléments de qualité des données	11
7.5 Éléments de métaqualité	14
7.6 Descripteurs d'un élément de métaqualité	15
8 Mesures de qualité des données	16
8.1 Généralités	16
8.2 Mesures de qualité des données normalisées	16
8.3 Mesures de qualité des données définies par l'utilisateur	16
8.4 Catalogue de mesures de qualité des données	16
8.5 Liste des composants	17
8.6 Détails des composants	18
9 Évaluation de la qualité de données	20
9.1 Processus d'évaluation de la qualité de données	20
9.2 Méthodes d'évaluation de la qualité de données	22
9.3 Agrégation et dérivation	23
10 Rendre compte de la qualité des données	24
10.1 Généralités	24
10.2 Cas particuliers	24
Annexe A (normative) Suite de tests abstraits	26
Annexe B (informative) Les concepts de la qualité des données et leur utilisation	28
Annexe C (normative) Dictionnaire de données pour la qualité des données	34
Annexe D (normative) Catalogue des mesures normalisées de la qualité de données	55
Annexe E (informative) Évaluation et rapportage de la qualité des données	106
Annexe F (informative) Méthodes par échantillonnage pour l'évaluation	131
Annexe G (normative) Mesures de base de la qualité des données	140
Annexe H (informative) Gestion des mesures de la qualité des données	145
Annexe I (informative) Lignes directrices relatives à l'utilisation des éléments de qualité	148
Annexe J (informative) Agrégation des résultats de la qualité des données	158
Bibliographie	160

$$\phi = \frac{1}{2} \arctan \frac{2\sigma_{x_k y_k}}{\sigma_{x_k}^2 - \sigma_{y_k}^2}$$

$$a = \sqrt{\frac{1}{2} \chi_{1-\alpha}^2(2) \left(\sigma_{x_k}^2 + \sigma_{y_k}^2 + \sqrt{(\sigma_{x_k}^2 - \sigma_{y_k}^2)^2 + 4\sigma_{x_k y_k}^2} \right)}$$

$$b = \sqrt{\frac{1}{2} \chi_{1-\alpha}^2(2) \left(\sigma_{x_k}^2 - \sigma_{y_k}^2 - \sqrt{(\sigma_{x_k}^2 - \sigma_{y_k}^2)^2 + 4\sigma_{x_k y_k}^2} \right)}$$



Une approche logicielle encore limitée

- Indispensable pour garantir l'usage des données et leur intégration dans des applicatifs ou des services
- Centrée sur le respect de structure et les règles topologiques (critère de cohérence logique au sens de la norme ISO 19157)
- Mais ne traite pas les aspects plus qualitatifs (exhaustivité, précision géométrique...)

Dit autrement: garantie d'avoir des résultats mais pas de garantie des résultats

Vers une approche plus pragmatique

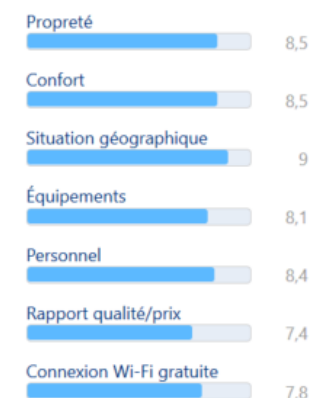
- Une simplification de la norme : sélection d'indicateurs pertinents et « aisément » calculables
 - L'acceptation d'approches moins scientifiques mais accessibles sur les plans technique et financier (le dire d'expert par exemple)
 - Des propositions de réponses adaptées aux différentes situations
 - Existence de spécifications
 - Existence de sources de contrôle pertinente
 - Néant
- Un ensemble de fiches méthodes (éléments statistiques, échantillonnage, méthodes de contrôle, modes de représentation...) associé à des fiches « critère qualité »

Une pensée pour les utilisateurs

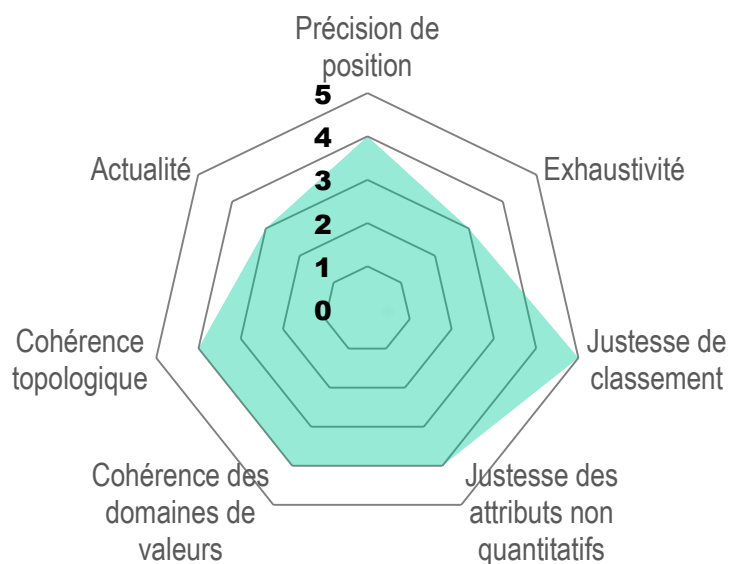
La qualité, la mesurer c'est bien, savoir en parler c'est mieux

Reporting à 3 niveaux

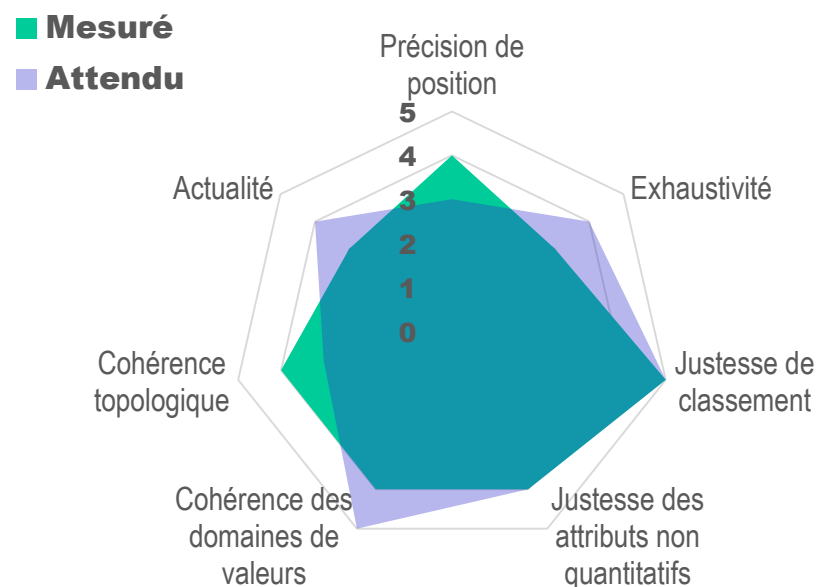
- Rapport qualité standard – le plus complet
 - Ensemble des mesures obtenues et méthodes employées.
 - Séparation des mesures par classes et/ou propriétés
- Rapport simplifié – solution intermédiaire
 - Ensemble des variables mesurées ramené à un nombre réduit d'indicateurs
 - Agrégation de mesures
 - Représentation graphique
- Représentation synthétique => une seule note



Représentation simplifiée



Sans spécifications



Avec spécifications






La difficulté : fixer les graduations

Exemple : précision de position

Indicateur: valeur moyenne des incertitudes

Note	Intervalle	commentaire
5	$m \leq 0,4 \text{ m}$	0,4 m correspond à la limite supérieure de la classe A définie dans l'arrêté des classes de précision (arrêté du 16 septembre 2003)
4	$0,4 \text{ m} < m \leq 1,5 \text{ m}$	1,5 m correspond à la limite supérieure de la classe B définie dans l'arrêté des classes de précision. Compatible avec une saisie sur le RGE
3	$1,5 \text{ m} < m \leq 5 \text{ m}$	Levés avec GPS de milieu de gamme
2	$5 \text{ m} < m \leq 20 \text{ m}$	Ce qu'il est possible d'obtenir en basant la saisie sur du Scan 25 ou de la BD Carto. C'est également l'ordre de grandeur de la précision d'un GPS standard en milieu urbain ou défavorable
1	$m > 20 \text{ m}$	
0	Sans objet	Critère non mesuré

Perspectives

- Prise en compte de l'usage dans la notation
 - Cartographie 
 - Traitement statistique 
 - Localisation 
 - Navigation 
 - Analyse spatiale 

Perspectives

- Considérer la « note de qualité » comme un critère de sélection dans les catalogues ou portails

The screenshot shows the 'Géocat catalogue' search interface. At the top, there are navigation links: 'Rechercher', 'Cataloguer', 'S'informer', 'Services', and 'Partenaires'. The search results are displayed for the query 'solaire'. The page shows 'Résultats 1 à 5 sur 46'. The search filters on the left include 'Thèmes INSPIRE' (Lieux de production et sites industriels (2)), 'Format' (ECW (Enhanced Compression Wavelet) (6), ECW (1), MapInfo TAB (1), TIFF (1), ZIP (1)), 'Type de ressource' (Données (42), Services (3), Cartes (1)), and 'Résolution spatiale' (1/25000 (10), 1/5000 (7)). The search results list includes:

- ★ **Potentiel solaire - Limousin**
 etude "potentiel solaire" en Limousin réalisée par le CETE Méditerranée. Plus d'informations : http://piece-jointe-carto.developpement-durable.gouv.fr/REG074B/SITE_INDUSTRIEL_PRODUCTION/N_SITE_PRODUCTION_ENERGIE/L_ETUDE_ENSOLEILLEMENT_R74/Notice_calcul_insolation.pdf...
 (5 stars rating)
- ★ **Énergies renouvelables : Potentiel solaire**
 -
 (4 stars rating)
- ★ **Parcs solaires du Cantal**
 Localisation des parcs photovoltaïques
 (5 stars rating)



Merci de votre attention

contacts

arnauld.gallais@cerema.fr
yves.bonin@cerema.fr

Cerema Méditerranée

24 Mars 2017 - Decryptageo