

## Connaître la qualité d'une donnée géographique fiabilise son utilisation

*La connaissance de la qualité des données, en sécurisant l'utilisateur, incite davantage à leur réutilisation.*

*Ce décryptage de la norme ISO 19157 a pour vocation de donner un cadre méthodologique pour qualifier les données lors de leur diffusion.*

**L'essor des données ouvertes et géolocalisées et la profusion d'usages existants et à venir nous rend tous progressivement producteur et utilisateur de données géographiques.**

**Les activités régaliennes ou les politiques publiques s'appuient sur de l'information maîtrisée où la qualité des données produites ou utilisées devient un entrant indispensable. Pour autant, tout le monde ne dispose pas des moyens des producteurs institutionnels de données et il paraît utile de fournir des recommandations et des méthodes plus adaptées au contexte de chacun, pour qualifier les données géographiques, communiquer sur les résultats obtenus voire savoir les interpréter. C'est l'objectif que s'est fixé le Cerema en proposant cette collection de fiches, à l'interface des productions et des usages.**

**Cette fiche est la première d'une collection complète relative à la qualification des données géographiques, permettant de mettre en pratique la norme ISO 19157 en des termes simples, notamment dans le cadre d'une démarche d'ouverture des données.**



# 1. Contexte et démarche

Les données géographiques sont le carburant indispensable au bon fonctionnement des systèmes d'information territoriaux. La connaissance de leur qualité est devenue une nécessité à l'heure de la Directive européenne Inspire et de l'ouverture des données publiques.

Les ré-utilisateurs des informations géographiques s'interrogent fréquemment sur la qualité des lots de données qu'ils téléchargent. En effet, les géo-catalogues proposent des métadonnées en général fort peu explicatives quant à la qualité des données.

Or, la méconnaissance de la qualité freine non seulement la mise à disposition des ressources existantes (le fournisseur ayant peur d'exposer une donnée imparfaite), mais également la réutilisation des données. En effet, le ré-utilisateur potentiel n'a pas les moyens de vérifier que la ressource ouverte identifiée dispose de la qualité suffisante pour l'usage souhaité.

Produire des données de qualité peut s'avérer coûteux. Ne pas les évaluer suivant une méthode reconnue, éprouvée et déterminée en fonction de leur réutilisation peut l'être davantage, voire entraîner des conséquences dramatiques par un usage de données de qualité insuffisante. Il ne s'agit donc pas de chercher à améliorer la qualité des données, mais de disposer d'une connaissance sincère de leur niveau de qualité, de les « qualifier ».

À cette fin, le Cerema a élaboré, dès 2005, un didacticiel sur les normes de qualification des données géographiques en vigueur à l'époque (ISO 19113, 19114 et 19138). Ce contenu reste aujourd'hui disponible sur le site du Cerema :

 [catalogue.territoires-ville.cerema.fr](http://catalogue.territoires-ville.cerema.fr)

Depuis, la norme ISO 19157 a agrégé les trois précédentes normes et fournit désormais un excellent référentiel pour la qualification des données géographiques. Cette norme reste néanmoins très technique et peu facilement abordable, bien que non obligatoire au regard du CEN<sup>1</sup> ou de l'AFNOR<sup>2</sup>.

Ceci a conduit le ministère du Développement Durable et le Cerema à produire une première ressource pédagogique sur le sujet :

 [http://formations-geomatiques.developpement-durable.gouv.fr/NAT009/ADL/Approche\\_qualite\\_des\\_donnees/co/70\\_Generation\\_web\\_Approche\\_qualite\\_des\\_donnees.html](http://formations-geomatiques.developpement-durable.gouv.fr/NAT009/ADL/Approche_qualite_des_donnees/co/70_Generation_web_Approche_qualite_des_donnees.html)

Le Cerema a ensuite développé une méthode simplifiée susceptible de qualifier les données conformément à la norme ISO 19157, dans le but de diminuer le coût et la complexité de l'évaluation de la qualité, tant pour le fournisseur que pour le ré-utilisateur. Cette méthode traduit et résume la norme ISO 19157 pour la rendre accessible à tous. Elle y apporte également des notions de représentation et de rapportage de la qualité, permettant ainsi aux différents acteurs de savoir quoi mesurer et comment restituer cette évaluation.

Elle se matérialise sous la forme d'un ensemble de fiches, présentées en partie 2 de ce document.

La partie 3 oriente la manière d'aborder les différentes fiches, proposant ainsi des parcours adaptés en fonction des attentes de chaque lecteur, au travers des fiches.

1 Comité européen pour la normalisation

2 Association Française pour la NORmalisation

## 2. Les fiches

Outre cette présente fiche « chapeau », le corpus méthodologique comprend actuellement 10 fiches réparties dans deux ensembles : les fiches « méthodes » et les fiches « critères ».

### Les fiches « méthodes »

Ces fiches rappellent les concepts et précisent les modes opératoires généraux.



Généralités sur la qualité des données géographiques

Cette fiche propose un rappel du contexte normatif, présente de manière synthétique les concepts généraux sur la qualité des données géographiques – critères, mesures et méthodes d'évaluation – et apporte des précisions sur différents aspects de la qualité.



Éléments de contexte pour le contrôle qualité

Connaître le cadre dans lequel est conduite l'évaluation de la qualité d'un jeu de données est indispensable pour déterminer les méthodes de contrôle à disposition. L'objet de cette fiche est de décrire les différentes situations que l'on peut rencontrer, en particulier en fonction des sources de contrôle à disposition, et d'émettre des préconisations générales sur les méthodes à mettre en œuvre.



Éléments statistiques

Il n'est généralement pas possible d'évaluer la qualité sur tous les objets d'un lot de données. Dans la plupart des cas, on va réduire les mesures à un extrait – ou échantillon – pour rendre le contrôle financièrement soutenable. Afin de fonder les mesures finales sur des procédés incontestables, il est important de connaître quelques éléments statistiques de base qui vont permettre d'étendre les résultats obtenus sur un extrait à l'ensemble du jeu de données.

Cette fiche apporte des éléments sur le dimensionnement des échantillons et des précisions sur des notions essentielles pour interpréter les résultats comme le niveau de confiance ou l'incertitude.



Méthodes d'échantillonnage

Quand il n'est pas possible de vérifier tous les objets d'un jeu de données, on procède à une extraction de quelques-uns sur lesquels portera l'évaluation. Cette fiche explicite les recommandations pour constituer des échantillons représentatifs en vue du contrôle qualité des données géographiques, ainsi que différentes méthodes d'échantillonnage spatial des données géographiques.



Mode de représentation

Évaluer la qualité d'un jeu de données est important. Mais l'effort reste vain si les résultats ne sont pas communiqués. Cette fiche propose trois méthodes de représentation, de la plus riche à la plus synthétique, pour partager une même vision de la qualité mesurée.

## Les fiches « critères »

Ces fiches décrivent et sélectionnent les critères et les mesures définis par la norme ISO 19157 et retenus ici pour l'évaluation de la qualité des données géographiques. Elles proposent également des méthodes d'évaluation et de rapportage de la qualité selon différents critères.



Cohérence logique

La cohérence logique se définit comme le degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit. De manière schématique, ce critère s'attache à qualifier « l'utilisabilité » technique des données.



Exhaustivité

Ce critère rassemble tout ce qui caractérise l'absence de données attendues ou leur présence à tort.



Précision thématique

La précision thématique se définit comme la précision des attributs quantitatifs, la justesse des attributs non quantitatifs et la justesse de classement. Ce critère s'adresse, non plus aux objets eux-mêmes, mais aux informations qu'ils portent.



Précision de position

La précision de position se définit comme la précision de la position des entités au sein d'un système de référence spatial. Elle concerne aussi bien la position d'une entité prise isolément que sa position par rapport aux objets qui l'entourent.



Qualité temporelle

Ce critère s'intéresse à la dimension temporelle des attributs d'objets d'une base de données géographiques. Il en vérifie la précision temporelle ainsi que la cohérence logique et chronologique.

## Structure des fiches

Les fiches critères présentent toutes le même plan afin d'en unifier les règles de lecture. Leur objectif est de :

- rappeler les définitions de la norme ;
- décrire les mesures à réaliser pour les critères et sous-critères ;
- préconiser des indicateurs et sous-critères ;
- détailler une (des) méthode(s) de contrôle ;
- préconiser un mode de représentation ou de notation pour le rapportage de la qualité ;
- résumer « ce qu'il convient de retenir ».

Les fiches « méthode » sont par nature d'une forme plus libre, mais font référence, lorsque nécessaire, à la norme ISO 19157.

### 3. Parcours des fiches suivant votre profil

Quel que soit le contexte, la lecture préalable de la fiche méthode sur les généralités de la qualité permet de fixer un cadre général et de donner les premiers repères qui seront utiles pour la suite.

Les fiches ont vocation à être utilisées selon le contexte. Il est proposé dans ce qui suit quelques parcours différenciés selon trois profils de lecteur donnés à titre indicatif.

#### ■ **Commander une production de données en incluant des exigences de qualité**

La lecture des fiches critères pertinentes au regard des données à produire et des usages envisagés permettra d'identifier les indicateurs à utiliser pour traduire les exigences de qualité et connaître les méthodes qui pourront être mises en œuvre pour s'assurer qu'elles sont respectées. La lecture de la fiche méthode sur le contexte du contrôle apportera un complément sur ce qu'il est possible d'exiger.

#### ■ **Comprendre à quoi correspondent les indications de qualité fournies avec des données**

La lecture de la fiche méthode sur la représentation de la qualité précisera à quoi correspondent les indications de qualité et comment elles sont obtenues.

Poursuivre par la lecture des fiches critères concernées permettra de mieux comprendre les méthodes employées pour obtenir les mesures de qualité. Enfin, la lecture de la fiche méthode sur l'échantillonnage pourra apporter des compléments utiles.

#### ■ **Mesurer la qualité d'un patrimoine ou d'une production de données en tant qu'administrateur et/ou producteur de données localisées**

Il convient de lire les différentes fiches méthode en commençant par celles sur le contexte du contrôle et l'échantillonnage avant d'aller plus loin avec les fiches critères qui vous concernent. Le parcours se termine avec la fiche méthode sur la représentation de la qualité.

## Série de fiches « Qualifier les données géographiques »

<b>Fiche n° 01</b>	<b>Connaître la qualité d'une donnée géographique fiabilise son utilisation</b>
<b>Fiche n° 02</b>	Généralités sur la qualité des données géographiques
<b>Fiche n° 03</b>	Éléments de contexte pour le contrôle qualité
<b>Fiche n° 04</b>	Éléments statistiques
<b>Fiche n° 05</b>	Méthodes d'échantillonnage
<b>Fiche n° 06</b>	Modes de représentation
<b>Fiche n° 07</b>	Critère de cohérence logique
<b>Fiche n° 08</b>	Critère d'exhaustivité
<b>Fiche n° 09</b>	Critère de précision thématique
<b>Fiche n° 10</b>	Critère de précision de position
<b>Fiche n° 11</b>	Critère de qualité temporelle

### **Contributeurs** ●●●

Fiche réalisée sous la coordination de Gilles Troispoux et Bernard Allouche (Cerema Territoires et ville)

#### **Rédacteurs**

Yves Bonin (Cerema Méditerranée), Arnaud Gallais (Cerema Ouest)

#### **Contributeurs**

Mathieu Rajerison, Silvio Rousic (Cerema Méditerranée)

#### **Relecteurs**

Benoît David (Mission information géographique MTES/CGDD), Stéphane Rolle (CRIGe PACA), Magali Carnino (DGAC), Stéphane Lévêque (Cerema Territoires et ville)

**Maquettage**  
Cerema Territoires et ville  
Service édition

**Impression**  
Jouve  
Mayenne

**Illustration couverture**  
© IGN 2017

Date de publication 2017  
ISSN : 2417-9701  
2017/55

© 2017 - Cerema  
La reproduction totale ou  
partielle du document doit  
être soumise à l'accord  
préalable du Cerema.

### **Contact** ●●●

accueil.dtectv@cerema.fr

**Boutique en ligne : [catalogue.territoires-ville.cerema.fr](http://catalogue.territoires-ville.cerema.fr)**

#### **La collection « Connaissances » du Cerema**

Cette collection présente l'état des connaissances à un moment donné et délivre de l'information sur un sujet, sans pour autant prétendre à l'exhaustivité. Elle offre une mise à jour des savoirs et pratiques professionnelles incluant de nouvelles approches techniques ou méthodologiques. Elle s'adresse à des professionnels souhaitant maintenir et approfondir leurs connaissances sur des domaines techniques en évolution constante. Les éléments présentés peuvent être considérés comme des préconisations, sans avoir le statut de références validées.

Aménagement et développement des territoires - Ville et stratégies urbaines - Transition énergétique et climat - Environnement et ressources naturelles - Prévention des risques - Bien-être et réduction des nuisances - Mobilité et transport - Infrastructures de transport - Habitat et bâtiment