

*Qualifier les données géographiques
Un décryptage de la norme ISO 19157*

Critère de cohérence logique

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 existant 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 propose des méthodes pour exprimer la qualité d'un jeu de données relatives à leur cohérence logique. Mesurer la cohérence logique n'a de sens que s'il existe des spécifications qui décrivent le schéma conceptuel et la structure physique attendus. La vérification se fait alors par rapport aux règles décrites.

Pour autant, même en l'absence de spécifications, certaines règles « de bon sens » peuvent être vérifiées comme la projection cartographique utilisée et l'unicité des identifiants d'objets.



1. Définition de la norme

La cohérence logique se définit comme le degré de cohérence interne des données selon des règles de modélisation et de spécification du produit. Elle se décompose en quatre sous-critères :

- **cohérence conceptuelle** : il s'agit du respect du schéma conceptuel des données ;
- **cohérence au domaine de valeurs** : il s'agit de l'appartenance des valeurs d'attributs aux plages de valeurs spécifiées ;
- **cohérence de format** : il s'agit de l'adéquation avec la structure physique du jeu de données ;
- **cohérence topologique** : il s'agit de l'exactitude des caractéristiques topologiques du jeu de données.

Remarque : ce critère englobe des sous-critères de natures variées puisqu'il s'intéresse autant à la partie descriptive des objets qu'à leur topologie et leur implémentation physique.

Ce critère ou ses sous-critères peuvent se mesurer pour chaque classe d'objets ou chaque attribut que l'on désire évaluer.

Les attributs de datation font l'objet d'un autre critère de qualité (qualité temporelle) et ne sont pas traités dans ce document.

2. Description des mesures à réaliser

La norme ISO 19157 propose un grand nombre de mesures pour les sous-critères définis précédemment. Dans un souci de clarté, seules les plus importantes sont présentées dans ce document.

2.1 Cohérence conceptuelle

Ce sous-critère identifie tous les écarts aux règles implicites ou explicites du schéma conceptuel. Ces règles peuvent être soit décrites dans les spécifications, soit implicites car correspondant à une compréhension commune du monde réel.

Exemple : la présence d'un aéroport dans un lac, même si l'impossibilité n'est pas spécifiée, peut apparaître comme un écart à une situation normale.

■ Non-conformité au schéma conceptuel

Définition ■ Indication qu'un élément ne se conforme pas aux règles du schéma conceptuel.

Type de valeur ■ Booléen (« true » indique qu'un élément **ne se conforme pas** aux règles du schéma conceptuel).

Exemple : absence d'une classe d'objets, non unicité d'identifiant.

■ Conformité au schéma conceptuel

Définition ■ Indication qu'un élément se conforme aux règles du schéma conceptuel.

Type de valeur ■ Booléen (« true » indique qu'un élément **se conforme** aux règles du schéma conceptuel).

■ Nombre d'éléments non conformes aux règles du schéma conceptuel

Définition ■ Comptage de tous les éléments du jeu de données qui ne se conforment pas aux règles du schéma conceptuel.

Type de valeur ■ Nombre entier.

■ Taux de non-conformité par rapport aux règles du schéma conceptuel

Définition ■ Nombre d'éléments du jeu de données qui ne sont pas conformes aux règles du schéma conceptuel par rapport au nombre total d'éléments devant figurer dans le jeu de données (selon le schéma conceptuel).

Type de valeur ■ Nombre réel (pourcentage).

■ Taux de conformité aux règles du schéma conceptuel

Définition ■ Nombre d'éléments du jeu de données conformes aux règles du schéma conceptuel par rapport au nombre total d'éléments.

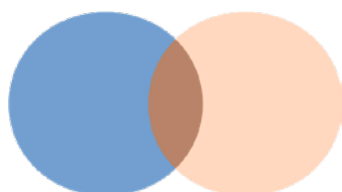
Type de valeur ■ Nombre réel (pourcentage).

Exemple : le pourcentage des objets « bâtiments » qui sont de géométrie polygonale.

■ Nombre de chevauchements de surfaces non valides

Définition ■ Comptage du nombre de chevauchements de surfaces non valides (« non valide » signifie que les objets des deux classes considérées ne devraient pas se chevaucher, selon les spécifications par exemple).

Type de valeur ■ Nombre entier.



Exemple : nombre de chevauchements entre les zones de dégagement aérien (servitude interdisant l'implantation d'éolienne) et les objets « éoliennes ».

Remarque : cette mesure, assez similaire à la détection des auto-chevauchements dans sa présentation, concerne des chevauchements non valides entre objets de classes différentes. Des chevauchements entre objets d'une même classe seront comptabilisés comme erreurs topologiques (cf. 2.4 « Nombre d'erreurs d'auto-chevauchement »).

2.2 Cohérence au domaine de valeurs

On entend par domaine de valeurs l'ensemble des valeurs (listes prédéfinies, intervalles) autorisées pour un attribut : cet ensemble est défini par une liste d'éléments énumérés ou une (ou plusieurs) condition(s) nécessaire(s) et suffisante(s) d'appartenance.

■ Non-conformité au domaine de valeurs

Définition ■ Indication qu'un élément n'est pas conforme à son domaine de valeurs.

Type de valeur ■ Booléen (« true » indique qu'un élément n'est pas conforme à son domaine de valeurs).

■ Conformité au domaine de valeurs

Définition ■ Indication qu'un élément est conforme à son domaine de valeurs.

Type de valeur ■ Booléen (« true » indique qu'un élément est conforme à son domaine de valeurs).

■ Nombre d'éléments non conformes à leur domaine de valeurs

Définition ■ Comptage de tous les éléments du jeu de données qui ne sont pas conformes à leur domaine de valeurs.

Type de valeur ■ Nombre entier.

■ Taux de conformité au domaine de valeurs

Définition ■ Nombre d'éléments du jeu de données qui sont conformes à leur domaine de valeurs par rapport au nombre total d'éléments du jeu de données.

Type de valeur ■ Nombre réel (pourcentage).

■ Taux de non-conformité au domaine de valeurs

Définition ■ Nombre d'éléments du jeu de données qui ne sont pas conformes à leur domaine de valeurs par rapport au nombre total d'éléments du jeu de données.

Type de valeur ■ Nombre réel (pourcentage).

Exemple : une base de données comporte une classe d'objets « franchissement » permettant de distinguer les franchissements entre routes et cours d'eau.

L'attribut « nature », de remplissage obligatoire, ne comprend que trois valeurs possibles : gué, pont, et tunnel. Sur les 3000 objets de la table « franchissement », une requête permet de détecter les valeurs aberrantes de cet attribut, à savoir : 12 valeurs « viaduc », 23 valeurs « passerelle », 14 objets non renseignés.

Ainsi, on peut renseigner les différentes mesures proposées par la norme ISO 19157 :

- non-conformité au domaine de valeurs : OUI
- conformité au domaine de valeurs : NON
- nombre d'éléments non conformes à leur domaine de valeurs : $12+23+14 = 49$
- taux de conformité au domaine de valeurs : $(3000-49)/3000 = 98\%$
- taux de non-conformité au domaine de valeurs : $49/3000 = 2\%$

Remarque : ne pas confondre (et compter deux fois) non-conformité au domaine de valeurs et précision sémantique (cf. fiche n°8 « Critère de précision thématique»). Par priorité, exclure les non-conformités au domaine de valeurs dans la population sur laquelle s'applique la précision sémantique.

Remarque : par souci de simplification et d'homogénéité, on intègre également dans la cohérence au domaine de valeurs, la validité temporelle et la cohérence temporelle, qui sont les éléments de qualité du critère, « Qualité temporelle » (voir fiche éponyme). En effet, la validité temporelle vérifie l'appartenance au domaine de valeurs temporelles (décrit comme l'ensemble des dates autorisées, les intervalles de temps, ou les durées dans lesquelles doivent obligatoirement s'inscrire les mesures temporelles), et la cohérence temporelle vérifie la cohérence chronologique entre objets d'une même classe ou entre les attributs d'un même objet. Les indicateurs, les méthodes et les conditions de contrôle sont similaires, notamment dans leur caractère automatisable à l'instar des contrôles relatifs au critère de cohérence logique.

2.3 Cohérence de format

La cohérence de format englobe les aspects liés au respect de la structure du lot de données, des noms de fichiers, des noms, du type, et de la taille des attributs, du stockage physique prévu, des formats de fichiers, etc.

Dans la norme, une différence par rapport au format des données est appelée « conflit de structure physique ».

■ Conflit de structure physique

Définition ■ Indication que le stockage des éléments entre en conflit avec la structure physique du jeu de données.

Type de valeur ■ Booléen (« true » indique un conflit de structure physique).

■ Nombre de conflits de structure physique

Définition ■ Comptage de tous les éléments du jeu de données dont le stockage entre en conflit avec la structure physique du jeu de données.

Type de valeur ■ Nombre entier.

■ Taux de conflits de structure physique

Définition ■ Nombre d'éléments du jeu de données dont le stockage entre en conflit avec la structure physique du jeu de données divisé par le nombre total d'éléments.

Type de valeur ■ Nombre réel (pourcentage).

Exemples :

- Tel fichier ne porte pas la dénomination attendue.
- La couche d'information est au format shapefile au lieu de GML.
- Un attribut INSEE est de longueur 3 caractères au lieu de 5.
- Le lot de données présente un encodage de caractères « ISO-8859-1 » au lieu de « UTF-8 ».
- Un pourcentage d'éléments ont un attribut date codé JJMMAAAA au lieu de AAAAMMJJ.

2.4 Cohérence topologique

Ce sous-critère regroupe plusieurs mesures caractérisant la qualité topologique de construction des objets géométriques (auto-intersections, micro-surfaces, connexions aux extrémités, etc.)

Il ne traite pas des relations topologiques entre deux classes d'objets (adjacence, inclusions, intersections) que l'on retrouve dans le § 2.1 sur la cohérence conceptuelle.

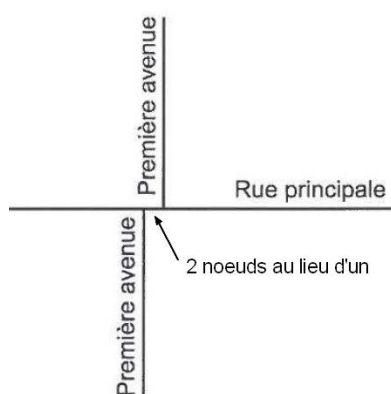
Il peut se contrôler avec des fonctionnalités dédiées des logiciels SIG.

■ Nombre de connexions arc-noeud erronées

Définition ■ Comptage des connexions arc-noeud erronées.

Type de valeur ■ Nombre entier.

Exemple : carrefour en « baïonnette » injustifié



■ Taux de connexions arc-noeud erronées

Définition ■ Nombre de connexions arc-noeud erronées par rapport au nombre de connexions supposées.

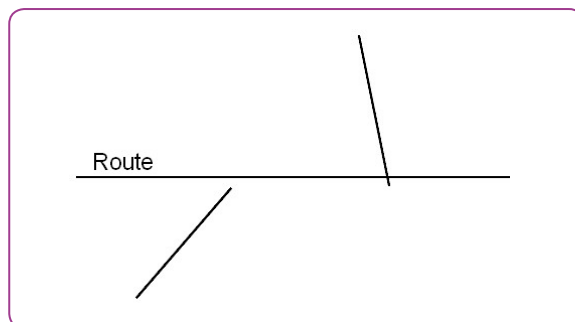
Type de valeur ■ Nombre réel (pourcentage).

■ Nombre de connexions manquantes en raison de ligne trop courte

Définition ■ Comptage des éléments du jeu de données disjoints en raison de ligne trop courte par rapport à une tolérance donnée.

Paramètre ■ Tolérance de recherche à partir de l'extrémité d'une ligne pendante.

Type de valeur ■ Nombre entier.



■ Nombre de connexions manquantes en raison de ligne trop longue

Définition ■ Comptage des éléments du jeu de données disjoints en raison de lignes trop longues.

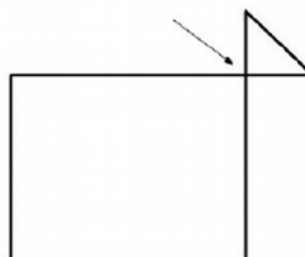
Paramètre ■ Tolérance de recherche sur la longueur minimale admise dans le jeu de données.

Type de valeur ■ Nombre entier.

■ Nombre de micro-surfaces non valides

Définition ■ Comptage de toutes les micro-surfaces non valides.

Description ■ Une micro-surface est une zone non souhaitée, apparaissant lorsque des surfaces adjacentes ne sont pas numérisées correctement. Les bords de surfaces adjacentes peuvent se chevaucher ou s'écarter de manière involontaire, en faible distance, et provoquer cette erreur topologique.



Paramètre 1 ■ Surface maximale d'une micro-surface (les éléments de surface supérieures ne seront pas considérés comme des micro-surfaces).

Paramètre 2 ■ Quotient d'épaisseur :

$Q = 4 \pi (\text{surface}) / (\text{périmètre})^2$: il est compris entre 0 (pour une ligne) et 1 (pour un cercle). Les éléments présentant un quotient supérieur à ce paramètre ne seront pas considérés comme des micro-surfaces.

Type de valeur ■ Nombre entier

■ Nombre d'erreurs d'auto-intersections non valides

Définition ■ Comptage de toutes les auto-intersections non valides.

Type de valeur ■ Nombre entier.



■ Nombre d'erreurs d'auto-chevauchement

Définition ■ Comptage de tous les éléments des données qui se chevauchent entre eux de façon illégitime. Cela concerne tout type de géométrie : doublon de points, chevauchements de surfaces, segment commun à deux arcs.

Type de valeur ■ Nombre entier.



Remarque : cette mesure est assez similaire à la détection du nombre de chevauchements de surfaces non valides (cf. 2.1). Elle ne concerne que les chevauchements entre objets d'une même classe.

3. Mesures retenues

On retiendra du chapitre précédent que les mesures possibles pour qualifier la cohérence logique sont à la fois nombreuses et variées puisqu'elles concernent des domaines aussi différents que la structure, le contenu et la topologie des lots de données.

Il convient donc d'en réduire le nombre pour isoler les mesures les plus pertinentes, tant dans leur compréhension que dans la capacité à les mesurer effectivement.

Pour le sous-critère **cohérence du domaine de valeurs**, il est proposé de retenir un taux de conformité aux domaines de valeurs, résultat de la moyenne des taux de conformités des attributs pour lesquels un domaine de valeurs existe. À noter qu'en présence de données temporelles, cette moyenne intègre les résultats obtenus pour la validité temporelle des attributs de datation ou d'horodatage.

Remarque : les attributs sans domaine de valeurs sont par définition hors du périmètre de ce critère qualité.

Pour le sous-critère **cohérence de format**, nous retiendrons le **conflit de structure physique**, sinon le jeu de données ne peut pas être exploité.

Concernant le sous-critère **cohérence topologique**, deux usages ressortent :

- pour la navigation, sur un réseau on retiendra le taux de connexions erronées ;
- pour les partitions surfaciques territoriales, on retiendra les trois mesures : nombre de micro-surfaces non valides, nombre d'erreurs d'auto-intersections non valides, nombre d'erreurs de chevauchement.

Remarque : par souci d'homogénéité et même s'ils ne sont pas prévus par la norme, on privilégiera des mesures sous forme de taux aux dénombrements.

Pour la **cohérence conceptuelle**, les anomalies que l'on peut identifier potentiellement et qui relèvent de cette catégorie concernent des points qui peuvent être très variables et difficiles à quantifier et à comparer, d'où la difficulté de ne ressortir qu'une seule mesure. Il est proposé, pour ce sujet

de ne pas essayer de mesurer par une valeur ou une note, ce sous-critère et de se contenter d'une description littérale des anomalies constatées qui viendront compléter les mesures des trois sous-critères précédents.

Exemple : contrôle qualité d'une base d'occupation du sol produite en régie. Sont retenues pour le critère cohérence logique les mesures :

- encodage des caractères en UTF8 (conflit de structure physique) => cohérence de format ;
- format des fichiers produits en shapefile (conflit de structure physique) => cohérence de format ;
- nommage / arborescence des fichiers (conflit de structure physique) => cohérence de format ;
- respect des listes de valeurs énumérées (taux de conformité au domaine de valeurs) => cohérence du domaine de valeurs ;
- taux de chevauchements (taux d'erreurs de chevauchements) => cohérence conceptuelle ;
- taux de « trous » (non prévu par la norme) => cohérence topologique ;
- taux de géométries invalides (taux d'erreurs d'auto-intersections non valides) => cohérence topologique ;
- taux d'objets zonage dont la surface est inférieure à la surface minimale de collecte (taux de micro-surface non valides) => cohérence topologique.

4. Méthode de contrôle

Mesurer la cohérence logique d'un jeu de données consiste à évaluer le degré de cohérence interne des données suivant les règles de modélisation et les règles inhérentes à la spécification de produit.

Ce critère touche, selon la norme ISO 19157, à des notions aussi variées que la cohérence conceptuelle, la cohérence aux plages de valeurs, au format, et la cohérence topologique.

Il concerne moins l'exactitude des informations que leurs cohérences internes logique et topologique, et n'implique donc pas d'avoir nécessairement recours à un jeu de données de référence.

Pour autant, plus que pour tout autre critère, les spécifications de produit sont ici nécessaires à l'évaluation du jeu de données : définition des classes d'objets, règles de sélection des objets, définition des plages de valeurs pour les attributs d'objets, règles topologiques, règles d'implémentation informatique, etc.

Les méthodes de contrôle de la cohérence logique diffèrent sensiblement de celles des autres critères sur deux points :

- il est possible de travailler sur l'ensemble du jeu de données et pas uniquement un échantillon ;
- le contrôle s'appuie sur des principes automatisables pour des jeux de données basés sur des mêmes spécifications.

Les méthodes prennent deux formes :

- analyse des différents fichiers composant le jeu de données et comparaison avec les attendus du schéma conceptuel et/ou de la description technique des fichiers : vérification de la structure, de l'emprise, du système de projection, présence d'un attribut identifiant... À cette étape, on ne s'intéresse pas aux objets eux-mêmes mais uniquement à leurs contenants : fichiers ou tables ;
- analyse du contenu des couches, soit par une synthèse statistique (vérification du respect

des domaines de valeurs ou des conditions d'appartenance, unicité des identifiants), soit par une analyse géométrique (géométries présentes, auto-intersections, détection des micro-surfaces, des chevauchements...). Cette analyse requiert l'utilisation de fonctions disponibles dans les différents outils SIG et ETL¹. Cette étape s'attache à qualifier le contenu.

4.1 Existence de spécifications de produit détaillées

Si l'on dispose de spécifications de produit détaillées, la méthode consiste à s'appuyer sur une batterie de tests, relativement automatisables, permettant de contrôler :

- l'adéquation au modèle conceptuel de données, y compris la cohérence aux plages de valeurs ;

- la cohérence avec les formats et les règles d'implémentation (arborescence, formats et noms de fichiers) ;
- la cohérence avec les règles topologiques spécifiées.

Cette batterie de tests peut être développée dans un outil de type valideur, dans lequel les différents contrôles à effectuer sont paramétrables sous forme de règles.

Exemple : le valideur INSPIRE² propose des tests détaillés pour aider les producteurs de données à connaître le niveau de conformité de leurs données avec les guides techniques INSPIRE.

Exemple : le valideur du Géoportail de l'urbanisme contrôle la conformité des documents d'urbanisme dématérialisés par rapport au standard CNIG déclaré dans les métadonnées.

Le valideur contrôle :

- le nom de l'archive zip encapsulant le jeu de données, son arborescence (répertoires et sous-répertoires), la présence des fichiers géographiques et pièces écrites obligatoires ;
- la présence de la fiche de métadonnée au format xml, l'encodage des caractères dans le jeu de données ;
- la présence et le nommage des attributs, le respect des plages de valeurs attributaires ;
- des anomalies topologiques élémentaires : objets de géométrie auto-intersectante, surface nulle.

Il ne contrôle pas :

- l'exhaustivité, ni la précision thématique des zonages (seul un humain pouvant le faire) ;
- des règles topologiques plus poussées comme le fait que le zonage d'urbanisme constitue une partition complète de territoire, que deux zonages ne peuvent pas se superposer, etc.

¹ ETL : outil d'import et de conversion de données (Extract Transform Load).

² Accès au valideur : <http://inspire-sandbox.jrc.ec.europa.eu/validator/>.

4.2 Absence de spécifications de produit détaillées

Si l'on ne dispose pas de spécification de produit, ou si elles sont incomplètes ou insuffisamment détaillées, ceci impactera particulièrement l'exploitation des sous-critères **cohérence conceptuelle** et **cohérence au domaine de valeurs**, tous deux évalués relativement au schéma conceptuel du jeu de données.

La méthode d'évaluation consistera dans ce cas à rechercher des informations potentiellement aberrantes dans le jeu de données, en particulier des valeurs d'attributs statistiquement exceptionnelles, susceptibles de se trouver en dehors des plages de valeurs communément admises.

Une autre possibilité consiste à s'appuyer sur un lot de données analogues faisant fonction de référence pour le schéma conceptuel attendu.

Exemple : un tronçon de route possédant 10 voies, ou un bâtiment de hauteur ou de géométrie inhabituelle seront à inspecter de près.

Pour le sous-critère « cohérence de format », certaines anomalies constatées (nom de table, de fichier ou d'attribut par exemple) peuvent être facilement corrigées sans ambiguïté, ce qui permet finalement d'exploiter le jeu de données sans recourir à une reprise de production.

5. Représentation – Notation

5.1 Représentation sous forme complexe

Cette représentation consiste à décrire l'ensemble des mesures effectuées et les résultats obtenus.

Elle demande un certain niveau de connaissances sur la conception et la structuration des bases de données et est à réserver prioritairement aux producteurs de données et aux commanditaires. Par ailleurs, et c'est encore plus vrai pour ce critère qualité, elle prend tout son sens en présence de spécifications, car elle permet de détailler l'analyse de la conformité d'un jeu de données.

5.2 Représentation sous forme simplifiée

Ce critère englobe des sous-critères de natures très différentes, il est donc particulièrement difficile de les agréger dans une notation unique. Ils sont donc traités de manière différenciée.

Il est proposé de retenir deux notes correspondant aux sous-critères quantitatifs (taux) : « cohérence des domaines de valeurs » et « cohérence topologique ».

■ Cohérence des domaines de valeurs

Cet élément de qualité est décrit par le taux de conformité aux domaines de valeurs, résultat de la moyenne des taux de conformités des attributs pour lesquels un domaine de valeurs existe (y compris attributs de datation ou d'horodatage).

La vérification de l'appartenance à un domaine de valeurs est un contrôle qui est la plupart du temps automatisable. Partant du principe que ce qu'il est possible de faire automatiquement sur l'ensemble du jeu de données lors de la phase d'évaluation qualité l'est également lors des phases de production à condition de mettre en place un processus qualité, les taux de conformité aux domaines de valeurs attendus sont élevés. D'autant qu'il est possible dans de nombreux cas de procéder aux corrections suite à ces contrôles internes.

Ainsi l'échelle de notes proposée est très volontariste.

Cohérence des domaines de valeurs taux de conformité aux domaines de valeurs	Note sur 5
De [99,5 % à 100 %]	5
De [99 % à 99,5 %]	4
De [97,5 % à 99 %]	3
De [95 % à 97,5 %]	2
Taux < 95 %	1

■ Cohérence topologique

Il s'agit ici de donner une échelle de graduation au taux de cohérence topologique dont la construction diffère selon que les données sont de type linéaire ou surfacique. Dans le premier cas, il reprend le taux de connexions erronées. Dans le second cas, il est issu de la moyenne des taux mesurés pour les auto-intersections, les micro-surfaces non valides et les auto-chevauchements, même si ces trois types d'erreurs topologiques n'ont pas les mêmes incidences sur les exploitations des données.

Comme pour la cohérence aux domaines de valeurs, ces erreurs topologiques sont détectables de manière automatique et sur l'ensemble du jeu de données. Il est donc possible de les détecter en amont dans le processus de production et, pour certaines d'entre elles, de les corriger automatiquement.

Cependant, pour tenir compte de la spécificité de la géométrie des objets et des problèmes rencontrés parfois dans les transformations de format, nous retenons une échelle de graduation ambitieuse certes mais moins exigeante de celle proposée pour les domaines de valeurs.

Cohérence topologique taux de cohérence topologique	Note sur 5
De [99 % à 100 %]	5
De [97,5 % à 99 %]	4
De [95 % à 97,5 %]	3
De [90 % à 95 %]	2
Taux < 90 %	1

Le sous-critère « **cohérence conceptuelle** » comportant de nombreux aspects très divers ne peut être réduit à une seule note d'autant que certaines anomalies ont des conséquences qui peuvent aller jusqu'à l'impossibilité d'utiliser un jeu de données. Pour autant, même si cet élément de qualité n'est pas traduit sous forme de notes, il est essentiel que la représentation détaillée puisse restituer les erreurs qui rendent le lot impropre à certaines exploitations. Aussi, ce sous-critère doit faire l'objet d'une liste des anomalies constatées, en distinguant les problèmes critiques (erreurs rédhibitoires) des avertissements (uniquement portés à titre indicatif), qui sera annexée au graphique de la représentation simplifiée.

Dans un même ordre d'idée, le sous-critère « cohérence de format » sera décrit de façon binaire par la présence ou non d'un conflit de structure physique. En cas de conflit de structure physique, cette information sera également fournie en annexe du graphique.

Ce qu'il faut retenir

Ce critère concerne moins l'exactitude des informations que leurs cohérences internes logique et topologique. Il n'implique donc pas d'avoir nécessairement recours à un jeu de données de référence.

Plus que pour tout autre critère, les spécifications de produit (spécifications de sélection d'objets, définition des plages de valeurs, etc.) sont nécessaires à l'évaluation de l'intégrité logique d'un jeu de données.

La cohérence logique est principalement utilisée pour le contrôle d'intégrité des bases de données et il s'agit d'un contrôle préliminaire essentiel.

En effet, le jeu de données ne peut pas être exploité pour certains usages ou certaines applications s'il ne correspond pas aux prescriptions d'implémentation physique requises (format, encodage, arborescence, etc).

Par ailleurs, une partie de ces contrôles peut être automatisée avec des outils spécifiques de type « validateur », notamment les contrôles portant sur la cohérence conceptuelle, cohérence de format et cohérence topologique, dans une opération portant sur l'ensemble du jeu de données.

Les mêmes opérations peuvent être réalisées avec des logiciels du marché offrant des possibilités étendues et paramétrables.

Les exigences liées à ces différents sous-critères sont plus fortes que pour les autres critères du fait même que les contrôles sont majoritairement automatisables et portent sur l'ensemble du jeu de données et qu'ils pourraient, par conséquent, être mis en œuvre en amont dans les processus de production.

Série de fiches « Qualifier les données géographiques »

Fiche n° 01	Connaître la qualité d'une donnée géographique fiabilise son utilisation
Fiche n° 02	Généralités sur la qualité des données géographiques
Fiche n° 03	Éléments de contexte pour le contrôle qualité
Fiche n° 04	Éléments statistiques
Fiche n° 05	Méthodes d'échantillonnage
Fiche n° 06	Modes de représentation
Fiche n° 07	Critère de cohérence logique
Fiche n° 08	Critère d'exhaustivité
Fiche n° 09	Critère de précision thématique
Fiche n° 10	Critère de précision de position
Fiche n° 11	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), Yvan Bédard (Professeur Honoraire à l'université Laval, CEO d'Intelli³).

Maquettage

Cerema Territoires et ville
Service édition

Impression

Jouve
Mayenne



Contact ●●●

accueil.dtectv@cerema.fr

Date de publication 2017
ISSN : 2417-9701
2017/61

Boutique en ligne : 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

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