

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

Critère de qualité temporelle

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 propose d'explicitier un des critères qualité de la norme ISO 19157 retenu dans cette collection de fiches : la qualité temporelle, qui s'intéresse à la dimension temporelle que peuvent porter certains objets dans une base de données. Plus précisément il s'attache à vérifier la précision des datations voire horodatages ainsi qu'à la cohérence logique et chronologique des dates. Ce critère n'a d'intérêt que pour des bases de données où les informations de datation représentent un réel enjeu dans l'exploitation des données.



1. Les définitions utilisées

La qualité temporelle est un critère de la norme ISO 19157. Il se définit comme la qualité des attributs temporels et des relations temporelles entre les entités.

Il se compose de trois sous-critères :

- **l'exactitude de la mesure temporelle** : exactitude des mesures temporelles décrites par rapport aux valeurs acceptées ou reconnues comme vraies ;
- **la cohérence temporelle** : justesse de la chronologie des événements ;
- **la validité temporelle** : validité des données en ce qui concerne les aspects temporels.

La mesure temporelle peut être soit un point défini dans le temps, soit une durée. Par mesure temporelle, on entend toute information de temps se composant d'un ou plusieurs des éléments suivants : jour ; semaine (de 1 à 52) ; mois ; année ; heure (minute, seconde) ; jour de la semaine.

Ces éléments peuvent être complétés d'informations précisant les références temporelles (calendrier, fuseau horaire, UTC, heure d'été et heure d'hiver).

Remarque : La conformité du format des dates et heures n'est pas évaluée ici et relève du critère de cohérence logique. Cf. fiche n° 07 « Critère de cohérence logique ».

On ne cherche pas à évaluer les informations temporelles des métadonnées mais bien les valeurs d'attributs d'objets correspondant à des mesures temporelles.

Exemple : date de l'arrêté d'une servitude d'utilité publique, heure de passage d'une ligne de transport collectif, année de construction d'un bâtiment, heures d'ouverture et fermeture d'un bâtiment public, etc.

Remarque : Dans les travaux antérieurs sur la qualité des données, apparaissait la notion d'actualité – ou de « fraîcheur » des données – qui n'est plus retenue dans la norme 19157 considérant que cette information relève plutôt du niveau des métadonnées d'un jeu de données et n'est plus un élément de qualité des données elles-mêmes.

2. Description des mesures à réaliser par sous-critère

2.1 Exactitude de la mesure temporelle

■ Exactitude temporelle avec un seuil de signification de 95 %

Définition ■ Demi-longueur de l'intervalle défini par deux limites inférieure et supérieure, dans laquelle la valeur véritable de l'instance temps se situe avec une probabilité de 95 %.

Exemple : Il y a 95 % de chance que l'évènement se soit produit entre 14h30 et 17h30. L'exactitude temporelle est égale à $(17h30 - 14h30) / 2$ soit : 90 mn.

Description ■ Dans le cas d'une quantité mesurée qui prend des valeurs réelles, il est impossible de donner la probabilité qu'une valeur unique soit la valeur véritable. Mais il est possible de donner la probabilité que la valeur véritable entre dans un certain intervalle. Cet intervalle est appelé

l'intervalle de confiance. Il est donné par la probabilité P que la valeur véritable se situe entre la limite inférieure et la limite supérieure. Cette probabilité P est également appelée seuil de signification.

Type de valeur ■ Mesure, nombre réel.

Variantes ■

- exactitude temporelle avec un seuil de signification de 50% ;
- exactitude temporelle avec un seuil de signification de 68,3% ;
- exactitude temporelle avec un seuil de signification de 99% ;
- exactitude temporelle avec un seuil de signification de 99,8%.

Remarque : cette mesure est similaire à la précision des attributs quantitatifs (cf. fiche n° 09 *Critère de précision thématique*) mais elle porte sur l'aspect particulier des datations.

2.2 Cohérence temporelle

La cohérence temporelle traduit une cohérence chronologique entre objets d'une même classe ou entre attributs d'une même objet.

Exemples :

- date de démolition nécessairement postérieure à la date de construction ;
- heure de passage à l'arrêt n°2 d'une ligne de bus nécessairement postérieure à l'heure de passage à l'arrêt n°1 ;
- date d'un arrêté d'approbation antérieure à la date de sa publication au journal officiel.

La chronologie peut se présenter comme les différentes versions d'un même objet (arrêté initial, arrêté modificatif...) et peut se traduire par plusieurs attributs d'un même objet (date de construction, date de démolition).

La cohérence temporelle n'a pas pour objectif de caractériser une erreur temporelle (cf. sous-critère exactitude de la mesure temporelle).

■ Erreur chronologique

Définition ■ Indication qu'un élément est ordonné de manière incorrecte par rapport aux autres éléments.

Type de valeur ■ Booléen. La valeur « true » indique que l'élément est ordonné de manière incorrecte.

Exemple : pour un objet « établissement recevant du public » la valeur de « heure d'ouverture » est postérieure à la valeur « heure de fermeture ». Dans ce cas, la valeur « true » indique qu'il y a une incohérence temporelle.

2.3 Validité temporelle

Le domaine de valeurs temporelles est l'ensemble des dates autorisées, un intervalle de temps, ou une durée dans lesquels doivent obligatoirement s'inscrire les mesures temporelles.

Exemple : les dates 29 février 2017 et 25/13/2014 ne sont pas valides. Une durée de parcours de 2 mn entre les deux terminus d'une ligne de bus est invalide.

Ce critère peut être traité avec les mêmes mesures que le sous-critère « cohérence au domaine de valeurs » du critère « cohérence logique » (cf fiche correspondante).

■ 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).

3. Mesures retenues

3.1 Exactitude de la mesure temporelle

De manière similaire aux choix déjà exposés pour la précision thématique ou la précision de position, on retient prioritairement **l'exactitude de la mesure temporelle avec un seuil de signification de 95 %** pour caractériser l'exactitude temporelle. Ce choix de 95 % de niveau de confiance s'appuie sur les éléments statistiques (cf. la fiche méthode n° 4 « Éléments statistiques » §6) où la vraie valeur recherchée est située dans un intervalle de confiance centré sur la moyenne des mesures avec un intervalle de plus ou moins deux fois l'écart-type (2σ).

Exemple : dans la base de données des chantiers en cours sur le territoire d'une collectivité territoriale, la date de début de travaux doit être connue au jour près dans la mesure où elle servira de référence pour faire courir des délais de réalisation.

L'exactitude de la mesure temporelle est toujours exprimée dans la même unité que l'attribut. Si l'attribut correspond à une date, l'exactitude sera donnée en jours. Si l'attribut correspond à une heure (par exemple heure de passage d'un transport collectif), l'exactitude sera indiquée en minutes.

Cette mesure présente de nombreuses similitudes avec la mesure retenue pour qualifier l'exactitude des attributs quantitatifs du critère « précision thématique ». Cela est vrai pour la mesure – même indicateur – et également pour les méthodes de contrôle où seul diffère le type d'informations à contrôler. Aussi, par souci de simplification et de cohérence, les attributs temporels seront assimilés à des attributs quantitatifs et tant leur mesure que leur restitution seront agrégées avec la précision thématique.

3.2 Cohérence temporelle

Pour la **cohérence temporelle**, la norme ne prévoit aucune mesure agréant le résultat d'un ensemble de mesures. Par analogie avec les choix faits sur l'exhaustivité ou la précision thématique, il est proposé de créer une mesure supplémentaire : le **taux de conformité chronologique** défini comme suit :

Définition ■ Nombre d'objets ordonnés de manière correcte par rapport aux autres éléments (ou dont les attributs sont ordonnés de manière correcte) par rapport au nombre total d'objets du jeu de données.

Type de valeur ■ Nombre réel (pourcentage).

Cette mesure présente l'intérêt d'être présentée sous forme de taux plus facilement interprétable tout en privilégiant une approche positive en quantifiant la conformité plutôt que la non-conformité.

3.3 Validité temporelle

Pour la validité temporelle, les différentes mesures proposées demandent un même effort pour le contrôle. Une approche positive des choses milite pour retenir préférentiellement **le taux de conformité au domaine de valeurs**.

4. Méthode de contrôle

Dans la directive Inspire, seul le thème « *Service d'utilité publique et services publics* » nécessite de contrôler la qualité temporelle, en faisant référence aux attributs des bâtiments des services publics tels que la date de mise en service ; la date de fin ; les heures de service (ouverture/fermeture) d'un service public.

Les informations temporelles sont néanmoins d'un usage beaucoup plus large dans les bases de données géographiques.

Quelques exemples :

- millésime : date de l'information associée à un objet ;
- mise à jour : liste des millésimes pour lesquels l'information concernant un objet est disponible ;
- fraîcheur : décalage entre le millésime lié à l'objet et sa date de livraison ;
- date de début, date de fin, date de création, date de publication, etc.
- heure, jour, semaine ;
- étendue temporelle, durée.

Comme indiqué ci-dessus, mesurer la qualité temporelle consiste à évaluer :

- l'exactitude temporelle avec un seuil de signification de 95 % ;
- la cohérence temporelle en déterminant le taux de conformité chronologique ;
- la validité temporelle en déterminant le taux de conformité au domaine de valeurs.

Remarque préalable sur le périmètre du contrôle :

les données temporelles peuvent faire référence à des attributs d'objets, à des classes d'objets ou encore à des thèmes de données dans leur intégralité. Certaines de ces informations sont par conséquent contenues dans les métadonnées et n'affectent pas les données elles-mêmes, notamment lorsqu'elles se rattachent aux classes d'objets ou aux thèmes de données. Dans le cadre de la qualification des données, nous nous limitons aux informations temporelles relatives aux objets et non à celles contenues dans les métadonnées.

Considérant les attributs dates comme des attributs quantitatifs, on peut contrôler l'exactitude temporelle en même temps et suivant les mêmes méthodes de contrôle de la précision thématique que les autres attributs.

De même, on peut traiter la validité temporelle en même temps que les autres attributs sur le respect des domaines de valeurs. Ce contrôle consiste à vérifier que le format des données est conforme et qu'aucune date n'est aberrante. Ceci sera soit réalisé visuellement, soit à l'aide d'une procédure automatique.

Ainsi, seule la cohérence temporelle utilisant le taux de conformité chronologique nécessite un contrôle spécifique. Celui-ci porte sur le respect de la chronologie des événements.

Exemple : dans un document d'urbanisme, la date de création d'un zonage est nécessairement antérieure à la date d'approbation du document. Ce qui se traduit en termes de valeurs d'attribut par la condition : $\text{DATAVALID} \leq \text{DATAPPRO}$

Comme dans l'exemple ci-dessus, ceci n'implique pas d'avoir recours à une base de données de référence. Par exemple si l'on veut contrôler la cohérence temporelle de l'ouverture et de la fermeture d'un bâtiment public, il suffit de contrôler que l'heure de fermeture soit postérieure à son heure d'ouverture. Cela n'exclut pas qu'en cas d'erreur détectée, les horaires soient corrects, mais ceci relève de l'exactitude de la mesure temporelle et non de sa cohérence.

Pour le contrôle de cohérence temporelle :

- il faut s'assurer que la base de données contient bien des informations temporelles et que certaines décrivent une chronologie ;
- le contrôle sera réalisé soit visuellement si les données sont peu nombreuses, soit à l'aide d'une procédure automatique soit par requête ;
- on calculera alors le taux de conformité chronologique.

5. Représentation - Notation

Comme pour les autres critères qualité, les mesures effectuées sur la dimension temporelle doivent pouvoir être restituées. Il faut cependant garder à l'esprit que les attributs de datation ou d'horodatage restent minoritaires dans les jeux de données usuels. Il convient par ailleurs de distinguer les attributs décrivant l'existence d'un objet des attributs apportant des informations spécifiques.

Exemple : la Directive Inspire prévoit pour de nombreux thèmes que tout objet soit renseigné sur son cycle de vie (`validFrom` et `validTo`). Certes, ces informations sont intéressantes, mais elles ne revêtent pas, la plupart du temps, la même importance qu'une date d'approbation pour des documents réglementaires opposables.

Ces deux éléments amènent à aborder la question de la représentation de manière légèrement différente, en particulier pour la représentation simplifiée.

5.1 Représentation sous forme complexe

Il s'agit ici de reporter dans le détail les différentes mesures effectuées. Ce mode de représentation s'adresse prioritairement aux producteurs de données et aux commanditaires et trouve tout son intérêt quand des spécifications existent (cf. fiche méthode n° 6 « *Modes de représentation* »).

La représentation complexe consiste à décrire sur quels éléments porte chaque mesure, de fournir les résultats obtenus en précisant le contexte du contrôle : source de référence quand elle existe, méthode de contrôle et taille de l'échantillon le cas échéant, objectifs de qualité s'ils étaient précisés dans des spécifications.

5.2 Représentation sous forme simplifiée

Contrairement aux quatre autres critères qui sont potentiellement évaluable sur tous les jeux de données, la qualité temporelle peut, en l'absence de toute donnée de datation, ne pas être à mesurer. Ainsi, il est opportun de ne pas sur-représenter un critère qui ne sera pas évalué systématiquement.

Il a été précisé précédemment que le sous-critère exactitude de la mesure temporelle pouvait être assimilé aux mesures d'exactitude des attributs quantitatifs (du critère précision thématique). Ainsi, quand ils sont évalués, les attributs de datation ou d'horodatage mesurés seront présentés en adoptant les méthodes de représentation des attributs quantitatifs.

De même, la validité temporelle qualifie l'appartenance des dates ou horaires à des domaines de valeurs, et la cohérence temporelle traduit les contraintes de cohérence logique entre les dates. Ces deux mesures peuvent donc être représentées en adoptant les méthodes de la cohérence au domaine de valeurs¹ dont la notation agrègera aussi bien les mesures obtenues pour les attributs classiques que pour les attributs temporels.

Ainsi, l'ensemble des représentations simplifiées du critère de qualité temporelle reprend les modes de représentation des autres critères.

¹ Cf. fiche n° 07 « *Critère de cohérence logique* » (cf exemple sur la cohérence chronologique entre les attributs `DATAVALID` et `DATAPPRO`).

Ce qu'il faut retenir

Le critère de qualité temporelle n'est pas le critère prépondérant dans l'évaluation d'une base de données géographiques, il est d'ailleurs rarement proposé dans les spécifications de qualité des thèmes Inspire.

Il recouvre trois sous-critères et onze mesures. Nous en retenons une par sous-critère :

- **l'exactitude de la mesure temporelle**, avec un seuil de signification de 95 % ;
- **la cohérence temporelle**, avec la mesure du taux de conformité chronologique. Il s'agit d'une mesure supplémentaire proposée par analogie avec les choix effectués pour l'exhaustivité et la précision thématique ;
- **la validité temporelle**, avec la mesure du taux de conformité au domaine de valeurs.

Si cette spécificité temporelle le distingue du critère de précision thématique, les mesures restent analogues.

Mis à part le sous-critère d'exactitude de la mesure temporelle, ce critère concerne moins l'exactitude des informations temporelles que leur cohérence interne, logique et chronologique.

En tant que contrôles portant sur l'aspect « sémantique » du lot de données, il est recommandé de réaliser les contrôles qualité des sous-critères cohérence et validité temporelle en même temps que ceux relatifs aux critères d'exhaustivité et de précision thématique.

Le contrôle de la qualité temporelle n'implique pas d'avoir nécessairement recours à un jeu de données de référence et une partie de ces contrôles peut être automatisée avec des outils spécifiques de type « validateur ». Les contrôles portant sur la cohérence et la validité temporelles sont notamment réalisables dans une opération portant sur l'ensemble du jeu de données.

Série de fiches « Qualifier les données géographiques »

Fiche n° 01	Connaitre 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), Arnauld 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
2017/65

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.