



Approche qualité des données

V1.31

18/08/2016



Légende



Entrée du glossaire



Abréviation



Référence Bibliographique



Référence générale



Table des matières

| | |
|---|-----------|
| Objectifs | 5 |
| Introduction | 7 |
| I - Approche qualité des données | 9 |
| II - Généralités | 11 |
| A. Les expressions de la qualité..... | 11 |
| B. Réflexion sur "qualité des données géographiques"..... | 12 |
| C. Les normes..... | 12 |
| III - Mise en œuvre | 15 |
| A. Critères de qualité ISO 19157..... | 15 |
| 1. Notions de base..... | 15 |
| 2. Exemples..... | 16 |
| B. Exercice : Critères de qualité..... | 22 |
| C. Contrôles à effectuer ISO 19157..... | 22 |
| D. Exercice : Normes..... | 23 |
| E. Mesures à effectuer ISO 19157..... | 24 |
| F. Exercice : Types de mesures à prendre en compte..... | 25 |
| G. Exercice : Norme ISO 19157..... | 26 |
| H. Exercice : Exemple de mesure..... | 26 |
| IV - Conclusion | 29 |
| Solution des exercices | 31 |



Objectifs

- Connaître les différents concepts liés à la qualité
- Comprendre la méthodologie de contrôle
- Être capable de spécifier la qualité dans une commande de données en précisant la méthode de contrôle
- Être capable de s'assurer de la qualité d'un lot de données préalablement spécifié



Introduction

Le temps à consacrer à la lecture de cette ressource est évalué à **1h45**. Cette ressource alterne des apports théoriques illustrés d'exemples et de quizz.

Il est particulièrement important de réaliser les quizz et de consulter leurs corrigés avec attention car ils vous aideront à vous familiariser avec les notions abordées.

Approche qualité des données

I

Pourquoi gérer la qualité des données géographiques ?

La mise en place de la fonction d'administration des données se doit d'inventorier le stock disponible et d'adapter le contenu du patrimoine à gérer aux besoins des utilisateurs, selon les types de données, leur statut, leur qualité et leur provenance.



Fondamental

La qualité des données est un autre critère de choix des données patrimoniales susceptibles d'être échangées à l'intérieur ou à l'extérieur du service. La description de la qualité des données géographiques doit permettre, d'une part, aux producteurs de vérifier dans quelle mesure leurs produits sont conformes à leurs spécifications et, d'autre part, aux utilisateurs de définir leurs exigences ou de connaître les possibilités et limites d'utilisation de ces produits.



| | |
|---|----|
| Les expressions de la qualité | 11 |
| Réflexion sur "qualité des données géographiques" | 12 |
| Les normes | 12 |

A. Les expressions de la qualité



Définition : La qualité interne

- Permet de mesurer la qualité d'un lot de données produit
- Elle nécessite des spécifications initiales qui décrivent précisément ce qui doit être produit (cahier des charges)
- Les méthodes de contrôle de la qualité interne sont décrites dans des normes ISO
- La qualité interne mesurée est alors l'écart entre les spécifications initiales et ce qui a été réellement produit



Définition : La qualité externe

- Elle se définit comme l'adéquation aux besoins
- Elle est peu utilisée aujourd'hui et ne dispose pas de méthode spécifique pour la mesurer
- Elle relève plus du domaine de la recherche



Remarque : La qualité perçue pour un usage spécifique

Attention cette qualité perçue est la perception d'un utilisateur pour un besoin spécifique.

- De plus en plus commun sur le web
- Système de notation pas encore utilisé pour les données géographiques
- Quelques études réalisées dans le domaine de la recherche



Attention

La méthode décrite ci-après relève exclusivement de la qualité interne.

B. Réflexion sur "qualité des données géographiques"

L'objectif de la qualité des données géographiques

Ce document introduit la notion de qualité des données géographiques et présente succinctement les différentes normes internationales qui sont aujourd'hui les seuls outils disponibles pour mesurer et diffuser la qualité des bases de données géographiques.

La complexité de ses outils réservés à un monde d'experts est inadaptée pour la plupart des utilisateurs.

Il dresse un bilan des dysfonctionnements constatés relatifs à la difficulté à mesurer et à communiquer les informations de qualité qui doivent impérativement accompagner toute donnée.

L'objectif principal de ce document est d'alerter la communauté géomatique afin d'initier un débat autour de cette problématique dans le but de trouver, ensemble, des solutions techniques et juridiques mieux adaptées.



Complément : Télécharger le document

Le document en téléchargement ci-dessous est né à la suite de nombreux échanges menés avec des personnes convaincues de l'importance de la qualité des données géographiques

Le document est téléchargeable sur <http://www.certu-catalogue.fr/>

C. Les normes

Les seuls outils et méthodes actuellement accessibles pour décrire la qualité interne des données géographiques sont disponibles dans la norme ISO 19157

Cliquez sur les maisons pour en savoir plus sur les normes correspondantes .

Les seuls outils et méthodes actuellement accessibles pour décrire la qualité des données géographiques sont disponibles dans plusieurs normes ISO.

La qualité des données géographiques est décrite dans plusieurs normes ISO qui sont les suivantes :

- ISO 19157 : elle comprenant trois orientations :
 - la description de différents critères et sous-critères nécessaires à la description de la qualité des données géographiques
 - la définition d'un cadre de méthodes permettant l'évaluation des critères définis et comment rapporter la méthode
 - la définition de différentes mesures spécifiques, essentiellement mathématiques et statistiques liées aux critères et sous-critères définit préalablement

Une norme liée plus indirectement à la qualité des données géographiques est :

- ISO 19115 : elle définit les métadonnées indispensables à la description des données.



Fondamental

Les spécifications des données d'Inspire pour la qualité font systématiquement référence à la norme 19157.

La norme ISO 19131 précise comment spécifier une base de données.

Les spécifications initiales permettent de préciser le contenu de la base de données à produire et décrivent les différents contrôles à mettre en place au cours de la production.

Une fois produite, la base de données est contrôlée à l'aide des méthodes et des critères définis dans les normes ISO. Les résultats obtenus sont alors comparés aux spécifications initiales. L'écart entre les deux constitue la qualité interne du lot de données ainsi produit.

La qualité, une fois mesurée, peut-être communiquée dans un rapport spécifique sur la qualité ou à l'aide de métadonnées qui décrivent chaque critère évalué.



| | |
|---|----|
| Critères de qualité ISO 19157 | 15 |
| Exercice : Critères de qualité | 22 |
| Contrôles à effectuer ISO 19157 | 22 |
| Exercice : Normes | 23 |
| Mesures à effectuer ISO 19157 | 24 |
| Exercice : Types de mesures à prendre en compte | 25 |
| Exercice : Norme ISO 19157 | 26 |
| Exercice : Exemple de mesure | 26 |

A. Critères de qualité ISO 19157

Nous allons voir ensemble quels sont les critères de qualités issus dans la norme ISO 19157

1. Notions de base



critère qualité

Les différents critères de la qualité

Les critères et sous-critères qualité définis dans la norme ISO 19157 sont les suivants :

- Exhaustivité : L'exhaustivité est la conformité de la présence ou de l'absence des éléments du jeu de données par rapport au terrain nominal.
 - Excédent : excédent de données dans le jeu de données
 - Omission : déficit de données dans le jeu de données
- Cohérence logique : degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit du jeu de données.

La cohérence logique a d'abord été utilisée pour le contrôle d'intégrité des bases de données. L'extension s'est ensuite faite en géographie pour contrôler la topologie et c'est aujourd'hui encore le principal contrôle. A ce contrôle topologique s'ajoute la notion de fiabilité générale des informations.

- Cohérence conceptuelle : cohérence avec les règles du schéma conceptuel des données.
- Cohérence des domaines de valeur : cohérence des valeurs avec valeurs logiques du domaine.
- Cohérence du format : degré avec lequel les données sont en adéquation avec la structure physique du jeu de données.
- Cohérence topologique : exactitude des caractéristiques topologiques explicitement encodées dans le jeu de données.
- Précision de position : elle est définie comme la précision de positionnement des données sur la Terre.
- Précision absolue : écart entre les valeurs des coordonnées saisies et des valeurs reconnues comme étant vraies.
- Précision relative : écart entre les valeurs des coordonnées saisies et celles des mêmes objets relatifs de référence reconnues comme étant vraies
- Précision géométrique de données matricielles : écart entre les valeurs des coordonnées des données maillées et des valeurs reconnues comme étant vraies.
- Précision thématique ou précision sémantique : La précision sémantique est la conformité des valeurs des éléments du jeu de données avec les valeurs de leurs homologues dans le terrain nominal. Elle porte sur la classification des objets, la codification des attributs et les relations entre objets.
- Justesse du classement : conformités de la classe attribuée aux éléments ou à leurs attributs.
- Justesse des attributs non-quantitatifs : capacité d'un attribut non-quantitatif à être juste ou faux.
- Précision des attributs quantitatifs : écart entre la valeur saisie d'un attribut quantitatif et sa valeur reconnue comme étant vraie.
- Qualité temporelle : qualité des attributs temporels et des relations temporelles entre objets.
- Exactitude de la mesure temporelle : écart des mesures de temps avec celles reconnues comme étant vraies.
- Cohérence temporelle : exactitude de l'ordre des événements.
- Validité temporelle : validité des données par rapport au format et au calendrier spécifié pour l'ensemble des données.
- Utilisabilité (usability) : degré de cohérence avec un ensemble spécifique d'exigences de qualité des données. Cet élément introduit la notion de « qualité externe » qui exprime plus l'adéquation à un besoin et qui est propre à chaque utilisateur. Il est décrit très succinctement dans la norme et n'offre aucune méthode pragmatique à son utilisation. En revanche, il ouvre un domaine particulièrement intéressant sur lequel il est urgent de travailler. Il est traduit dans la directive Inspire comme la possibilité offerte aux utilisateurs qui auraient déjà travaillé avec une base de données de décrire les usages qu'elle a permis.

2. Exemples

a) Exhaustivité

L'exhaustivité est la conformité de la présence ou de l'absence des éléments du jeu de données par rapport au terrain nominal.

Les sous critères liés au critère exhaustivité sont excédent et omission



Exemple : Exemple pour le sous-critère excédent :

Dans un département, on désire vérifier sur une base de données dédiée aux établissements d'enseignement si le nombre d'écoles primaires est exhaustif.

La liste de référence publiée par le ministère de l'enseignement en recense 372. On en compte 383 dans la base de données contrôlée.

Nombre d'éléments en excès : 11

Taux d'éléments en excès : $11/372 = 3\%$



Exemple : Exemple pour le sous-critère omission :

Dans la même base de données, on désire vérifier les lycées publics.

La liste de référence publiée par le ministère de l'enseignement en recense 28. On en compte 26 dans la base de données contrôlée.

Nombre d'éléments manquants : 2

Taux d'éléments manquants : $2/28 = 7\%$

b) Cohérence logique

La cohérence logique est le degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit du jeu de données.

Les sous critères liés au critère cohérence logique sont cohérence conceptuelle, cohérence des domaines de valeur, cohérence du format, cohérence topologique.



Exemple : Le sous-critère cohérence conceptuelle :

Le contrôle peut se faire avec des outils spécifiques.



Exemple : Le sous-critère cohérence du format :

La cohérence de format englobe les aspects liés au respect des noms de fichiers, noms d'attributs, stockages physique prévu (nombre de caractères, type), format de fichiers... Dans la norme, une différence par rapport au format des données est appelé conflit de structure physique. Ce type de contrôle est assez aisément réalisé avec des outils spécifiques.



Exemple : Le sous-critère cohérence des domaines de valeur :

Une base de données topographique comporte une classe d'objets "franchissement" qui permet d'identifier les divers franchissement d'une route et d'un cours d'eau.

L'attribut "nature" ne comprend que trois valeurs possibles : gué, pont, et tunnel.
Sur les trois mille objet 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 donc renseigner les différentes mesures proposées par la norme ISO 19157 :

Non conformité au domaine de valeur : OUI

Conformité au domaine de valeur : NON

Nombre d'éléments non conformes à leur domaine de valeur : 49

Taux de conformité au domaine de valeur : $2951/3000 = 98\%$

Taux de non conformité au domaine de valeur : $51/3000 = 2\%$



Exemple : Le sous critère cohérence topologique :

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

Il se contrôle également avec des outils spécifiques.

c) Précision de position

La précision de position est définie comme la précision de positionnement des données sur la terre.

Les sous critères liés au critère précision de position sont précision absolue, précision relative et précision de position de données matricielles.



Exemple : Le sous-critère précision de position

On désire connaître la précision géométrique d'une base de données qui a été numérisée assez grossièrement sur une carte papier.

Afin d'estimer cette précision, on la superpose avec la BD TOPO de l'IGN dont on connaît sa précision métrique et on mesure les écarts en X et en Y d'une vingtaine de points homologues sur les deux bases.

Δx mesurés (en mètres) : 12, 6, 8, 9, 13, 10, 7, 12, 11, 9, 8, 11, 12, 9, 6, 7, 8, 10, 9, 12

Moyenne : 9,45 m Écart-type : 2,14 m

Δy mesurés (en mètres) : 8, 9, 10, 12, 14, 13, 8, 9, 10, 11, 12, 10, 9, 8, 7, 10, 13, 11, 9, 6

Moyenne : 9,95 m Écart-type : 2,11 m

Les mesures obtenues montrent que sur 20 points, les moyennes des écarts sont de 9,45 m en X et 9,95 m en Y avec des écarts-types de l'ordre de 2 m.

On peut donc en déduire que la précision moyenne est de l'ordre de 10 m par rapport à la BD TOPO.

La précision de la BD TOPO étant métrique, celle de notre base évaluée est alors déca-métrique.



Exemple : Le sous-critère précision relative

On désire connaître l'écart de précision entre deux bases dont on ne connaît pas la précision géométrique. On superpose alors la géométrie des deux bases et on mesure

les écarts en X et en Y d'une vingtaine de points homologues sur les deux bases.

En imaginant que les mêmes valeurs numériques soient les mêmes celles de l'exemple précédent, on pourrait juste en conclure que l'écart de précision entre les deux bases est de l'ordre de la dizaine de mètres sans connaître pour autant la précision absolue de chacune des deux bases.



Exemple : Le sous-critère précision de données matricielles :

Exemple d'une orthophotographie :

En règle générale, la précision d'une orthophotographie est le double de sa résolution.

Par exemple, pour la BD ORTHO de l'IGN dont la résolution est de 50 cm, on peut donc en déduire que sa précision géométrique est de l'ordre du mètre.

Exemple d'une carte scannée :

On estime la précision d'une carte papier à environ deux dixièmes de mm mesurés sur cette carte à l'échelle de la carte.

Par exemple, pour une carte au 1/10 000, deux dixièmes de mm à cette échelle représentent 2 mètres.

La précision est alors estimée à 2 mètres ce qui demeure un ordre de grandeur.

Pour une carte scannée, le scanner utilisé et la méthode employée, selon que l'on numérise une carte papier ou un support plus stable peuvent introduire une imprécision supplémentaire.

Dans notre exemple de la carte au 1/10 000, la précision est alors au mieux 2 m mais est très certainement supérieure.

D'une façon générale, pour évaluer la précision de données matricielles, on pourra les superposer avec une base de donnée vecteur de référence et appliquer la méthode expliquée pour la précision absolue.

d) Précision thématique

La précision thématique est la conformité des valeurs des éléments du jeu de données avec les valeurs de leurs homologues dans le terrain nominal.

Les sous critères liés au critère précision thématique sont la justesse du classement, la justesse des attributs non quantitatifs, la précision des attributs quantitatifs.



Exemple : Le sous-critère Justesse des attributs non quantitatifs

Un parc naturel recense la totalité de ses arbres dans une base de données géographiques. La classe d'objets des arbres possède l'attribut « essence ».

Un contrôle exhaustif terrain est effectué pour vérifier si l'essence noté dans la base est correcte. Le parc naturel compte 2800 arbres dans son emprise.

On dénombre, après contrôle, 35 erreurs dans la base où les arbres étaient renseignés avec une essence incorrecte.

Nombre de valeurs d'attribut incorrectes : 35

Taux de valeurs d'attribut correct : $2765/2800 = 98,75 \%$

Taux de valeurs d'attribut incorrectes : $35/2800 = 1,25 \%$



Exemple : Le sous-critère Justesse du classement

Une base de données qui regroupe les différents bâtiments d'un territoire différencie les bâtiments privés des bâtiments publics selon deux classes d'objets distinctes.

On désire contrôler la qualité de cette classification. On prélève ainsi un échantillon de 200 bâtiments dans la base et un contrôle terrain est effectué.

Sur les 200 bâtiments contrôlés 168 sont des bâtiments privés et 32 des bâtiments publics.

Les contrôles effectués ont permis de détecter 8 erreurs : 6 bâtiments privés étaient codés en public, et 2 bâtiments publics étaient codés en privé.

Nombre d'entités classées de manière incorrecte : 8 (6 bâtiments privés et 2 bâtiments publics)

Taux de classement erroné : $8/200 = 4 \%$

Taux de classement erroné pour les bâtiments privés : $6/168 = 4 \%$

Taux de classement erroné pour les bâtiments publics : $2/32 = 6 \%$

On peut représenter le résultat obtenu par la matrice de confusion suivante :

| | Base de données | |
|-------------------|-----------------|--------------|
| | Bât. privés | Bât. publics |
| Contrôle terrain | | |
| Bât. Privés : 168 | 162 | 6 |
| Bât. Publics : 32 | 2 | 30 |



Exemple : Le sous-critère Justesse des attributs quantitatifs

Une base de données sur l'habitat collectif recense l'ensemble des bâtiments d'habitation collective d'une ville. Un attribut de la classe d'objets « bâtiment » recueille le nombre estimé d'habitants par bâtiment.

On désire contrôler la méthode utilisée pour l'estimation du nombre d'habitants par bâtiment.

Un échantillon de 20 bâtiments est choisi et un sondage précis est réalisé afin de compter de façon exhaustive le nombre d'habitants.

| Bâtiments | Nombre réel d'habitants après sondage | Nombre évalué d'habitants dans la base de données | Delta |
|-----------|---------------------------------------|---|-------|
| 1 | 172 | 168 | 4 |
| 2 | 224 | 236 | 12 |
| 3 | 120 | 132 | 12 |
| 4 | 135 | 130 | 5 |
| 5 | 96 | 102 | 6 |
| 6 | 58 | 54 | 4 |
| 7 | 156 | 168 | 12 |
| 8 | 214 | 224 | 10 |

| | | Mise en œuvre | |
|----|-----|---------------|-----|
| 9 | 112 | 110 | 2 |
| 10 | 86 | 82 | 4 |
| 11 | 74 | 78 | 4 |
| 12 | 62 | 68 | 6 |
| 13 | 152 | 165 | 13 |
| 14 | 143 | 140 | 3 |
| 15 | 76 | 86 | 10 |
| 16 | 38 | 42 | 4 |
| 17 | 89 | 92 | 3 |
| 18 | 48 | 51 | 3 |
| 19 | 248 | 259 | 11 |
| 20 | 184 | 192 | 12 |
| | | Total | 140 |

Moyenne des écarts : $140/20 = 7$ habitants

Écart-type : $3,93$ habitants ~ 4

On peut donc estimer, avec un intervalle de confiance de 95 %, que le nombre d'habitants évalué dans la base de données est égal au nombre $N \pm 8$ ($\pm 2\sigma$).

e) Qualité temporelle

La qualité temporelle est la qualité des attributs temporelles et des relations temporelles entre les objets.

Les sous critères liés au critère qualité temporelle sont l'exactitude de la mesure temporelle, la cohérence temporelle et la validité temporelle



Exemple : Le sous-critère exactitude de la mesure temporelle

Cette mesure est similaire à celle des attributs quantitatifs (voir l'exemple du sous critère précision des attributs quantitatifs))



Exemple : Le sous-critère cohérence temporelle

Exemple 1 :

Une base de données des services publics d'une ville comporte, pour chaque service public, les attributs ; « heure d'ouverture » et « heure de fermeture ».

On désire contrôler sa cohérence temporelle, c'est-à-dire que l'heure de fermeture soit postérieure à l'heure d'ouverture. Il faut donc pour chaque service saisi dans la base, soit le vérifier visuellement, soit le vérifier à l'aide d'une requête ou d'une application spécifique.

Exemple 2 :

Le standard de données COVADIS sur le domaine éolien terrestre comporte la classe d'objet « éolienne » qui comprend les trois attributs temporels suivants : « date de

construction », « date de mise en production » et « date de démantèlement ». Ces trois dates sont liées chronologiquement, sachant que la construction a lieu avant la mise en production et que le démantèlement est la dernière phase.

Contrôler leur cohérence temporelle consiste à vérifier l'ordre logique de ces trois dates pour chaque éolienne.



Exemple : Le sous-critère validité temporelle

Toujours dans le standard de données COVADIS sur le domaine éolien terrestre, outre les trois dates citées dans l'exemple précédent, la classe « éolienne » comporte également un attribut de type booléen qui permet de savoir si l'éolienne est en service ou non.

On contrôlera alors la validité temporelle de l'attribut « en service » en vérifiant que pour toutes les valeurs OUI, la date de l'attribut « date de mise en service » est bien remplie et que pour la valeur NON, aucune date ne figure dans cet attribut.

B. Exercice : Critères de qualité

[Solution n°1 p 31]

Sélectionnez les définitions et déplacez les sous les mots leurs correspondants.

- i - C'est la qualité des attributs temporels et des relations temporelles entre objets
- ii - C'est le degré de cohérence avec un ensemble spécifique d'exigences de qualité des données
- iii - C'est la conformité des valeurs des éléments du jeu de données avec les valeurs de leurs homologues dans le terrain nominal
- iv - C'est le degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit du jeu de données
- v - C'est la conformité de la présence ou de l'absence des éléments du jeu de données par rapport au terrain nominal
- vi - C'est la précision de positionnement des données sur la Terre

| | | | | | |
|--------------|-----------|--------------|--------------|------------|---------------|
| Exhaustivité | Cohérence | Précision de | Précision | Qualité | Utilisabilité |
| : | logique : | position | thématique | temporelle | |
| | | | ou précision | | |
| | | | sémantique | | |

C. Contrôles à effectuer ISO 19157



Fondamental

La méthode de contrôle est déterminée essentiellement par la volumétrie des données à produire.

Il est donc important de déterminer si le contrôle est exhaustif ou s'il est effectué par échantillonnage qui reste le cas le plus fréquent.

Cette dernière méthode doit être suffisamment décrite pour permettre de réitérer des contrôles lors de mises à jour ou lors de contre-expertises.

Elle doit a minima définir :

- la nature de l'échantillon (couverture, nombre d'occurrences...)
- la nature du contrôle (méthode, système référent - source de contrôle...)
- le résultat du contrôle (avec son unité)

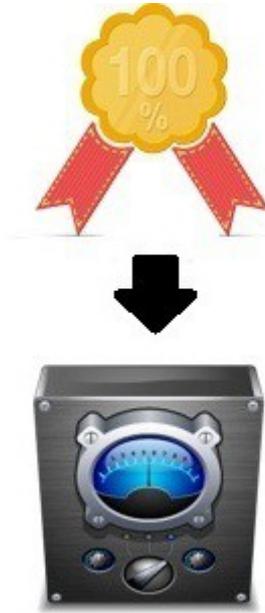


Image 1 critère controle



Méthode

Pour être complète la grille de contrôle doit donc comporter :

- ce sur quoi porte le contrôle
- ce que l'on mesure
- comment le mesure-t-on
- quelle est l'expression attendue du résultat final

D. Exercice : Normes

[Solution n°2 p 32]

Sélectionner parmi les phrases suivantes les 3 affirmations correctes

| | |
|--------------------------|--|
| <input type="checkbox"/> | Les différents critères et sous-critères définis dans la norme ISO 19157 ne doivent pas être systématiquement utilisés dans tous les cas |
| <input type="checkbox"/> | La norme ISO 19157 est meilleure que la norme 19115 |
| <input type="checkbox"/> | Le contenu et la spécificité de la base de données permettent de savoir quels sont les critères les mieux adaptés pour en contrôler la qualité |
| <input type="checkbox"/> | Il est important de connaître les différents critères proposés pour assurer leur adéquation avec les différents types de contrôles à mettre en place |
| <input type="checkbox"/> | La mise en pratique des normes se fait uniquement par les experts |

E. Mesures à effectuer ISO 19157

La norme ISO 19157 propose les différentes mesures qu'il est possible de réaliser par critère. Plusieurs mesures sont proposées par critère. On choisit la ou les mesures les mieux adaptées en fonction du type de donnée à évaluer.

Chaque mesure prévue doit être parfaitement décrite. Les paramètres suivants aident à cette description.

- Le champ d'application de la mesure : (c'est à dire ce que l'on s'engage à mesurer), il s'agit de l'échantillon prélevé s'il existe en incluant :
 - sa composition : type d'objet : point, ligne, polygone...
 - le nombre d'occurrences de cet échantillon : nombre d'objets concernés par le contrôle ;
 - sa situation géographique (emprise et surface) ;
 - et la date concernée par la mesure
 - exemple : évaluation du nombre d'auto-intersections concernant l'ensemble des objets surfaciques de la commune X.
- Le type de mesure effectuée :

C'est la description de la formule présentes dans le jeu de données cible et dans le jeu de contrôle

Exemple : ratio entre le nombre d'auto-intersections trouvés sur les objets surfaciques de la commune X et le nombre total d'objets surfaciques de cette commune.
- La procédure utilisée pour la mesure :

C'est à dire la méthode employée pour détecter les anomalies éventuelles

Cela peut-être :

 - un contrôle visuel par superposition à l'écran ;
 - un contrôle visuel par tirage papier (ou calque) ;
 - détection par requête spatiale (pour détecter les non-conformités de technologie par exemple) ;
 - par requête dans la base de données ;
 - ...

Le résultat obtenu

 - Il s'agit du résultat de la mesure spécifiant le type valeur et l'unité. Exemple : 95,7% Exemple : requête sql permettant de compter le nombre total d'objets surfaciques dans la commune X et le nombre d'objets affectés par l'erreur d'auto-intersection.

- Le type de valeur de l'unité :
Il s'agit de l'expression du résultat de la mesure de qualité distance, ratio, pourcentage, variable booléenne... Il est déterminé en fonction du mode de calcul.
Y inscrire aussi l'unité (nombre entier, réel à une décimale...) pour parfaire sa description.
Exemple : pas vraiment nécessaire car déjà précisé (nombre entier, réel à une décimale, pourcentage...)
- La date du contrôle qualité effectué sur l'indicateur
Date à laquelle est effectué le contrôle
Exemple : la aussi, on sait ce qu'est une date

...

Exemple de mesures

Pour chaque critères à évaluer, la norme propose un certain nombre de mesures.

- Indicateur d'erreur (booléen : la valeur est juste ou fausse)
- Nombre d'erreurs : nombre total d'erreurs mesurées sur l'échantillon
- Moyenne
- Écart type
- Taux d'erreur : pourcentage d'erreurs relevées sur l'échantillon évalué
- Matrice de co-variance
- ...



Image 2 critère contrôle mesure

F. Exercice : Types de mesures à prendre en compte

[Solution n°3 p 32]

Sélectionnez les explications des types de mesures et déplacez les sous types de mesures à effectuer.

- i - C'est la date à laquelle est effectuée le contrôle
- ii - C'est la description de la formule utilisée
- iii - C'est ce que l'on s'engage à mesurer

- iv - C'est l'expression du résultat de la mesure de qualité
- v - C'est la méthode employée pour détecter les anomalies éventuelles

| Le champ d'application de la mesure | Le type de mesure effectuée | La procédure utilisée pour la mesure | Le type de valeur et l'unité | La date du contrôle |
|-------------------------------------|-----------------------------|--------------------------------------|------------------------------|---------------------|
|-------------------------------------|-----------------------------|--------------------------------------|------------------------------|---------------------|

G. Exercice : Norme ISO 19157

[Solution n°4 p 33]

Sélectionner parmi les phrases suivantes les 2 affirmations correctes

| | |
|--------------------------|--|
| <input type="checkbox"/> | La norme ISO 19157 est meilleure que la norme 19115 |
| <input type="checkbox"/> | La norme ISO 19157 propose les différentes mesures qu'il est possible de réaliser par critère |
| <input type="checkbox"/> | On choisit la ou les mesures les mieux adaptées en fonction du type de donnée à évaluer. Certaines mesures sont très simples comme un simple comptage, une moyenne... D' autres sont plus complexes et nécessitent des mesures statistiques. |
| <input type="checkbox"/> | Il est important de connaître les différents critères proposés pour assurer leur adéquation avec les différents types de contrôle. On choisit la ou les mesures les mieux adaptées en fonction du type de donnée à évaluer. Pour chaque type de mesure, on applique des formules statistiques très compliquées |
| <input type="checkbox"/> | La mise en pratique de la norme 19157 se fait uniquement par les experts en statistique |

H. Exercice : Exemple de mesure

[Solution n°5 p 34]

Pour chaque critères à évaluer la norme propose un certain nombre de mesures

| | |
|--------------------------|---|
| <input type="checkbox"/> | Mise en œuvre |
| <input type="checkbox"/> | Indicateur d'erreur (booléen : la valeur est juste ou fausse) |
| <input type="checkbox"/> | Nombre d'erreurs : nombre total d'erreurs mesurées sur l'échantillon |
| <input type="checkbox"/> | Taille |
| <input type="checkbox"/> | Poids |
| <input type="checkbox"/> | Taux d'erreur : pourcentage d'erreurs relevées sur l'échantillon évalué |
| <input type="checkbox"/> | Matrice de co-variance |
| <input type="checkbox"/> | |



Attention

Une fois effectués, les résultats des différents contrôles qualité doivent être consignés soit dans des métadonnées, soit dans un rapport spécifique à la qualité des données.

BRAVO !

Vous connaissez à présent les différents concepts liés à la qualité.

Vous avez compris la méthodologie de contrôle

Vous êtes capable de vous assurer de la qualité d'un lot de données préalablement spécifié



Image 3 critère controle mesure



Solution des exercices

> Solution n°1 (exercice p. 22)

| | |
|--|---|
| Exhaustivité : | C'est la conformité de la présence ou de l'absence des éléments du jeu de données par rapport au terrain nominal |
| Cohérence logique : | C'est le degré de cohérence interne des données selon des règles de modélisation et les règles inhérentes à la spécification de produit du jeu de données |
| Précision de position | C'est la précision de positionnement des données sur la Terre |
| Précision thématique ou précision sémantique | C'est la conformité des valeurs des éléments du jeu de données avec les valeurs de leurs homologues dans le terrain nominal |
| Qualité temporelle | C'est la qualité des attributs temporels et des relations temporelles entre objets |
| Utilisabilité | C'est le degré de cohérence avec un ensemble spécifique d'exigences de qualité des données |

> Solution n°2 (exercice p. 23)

| | |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | Les différents critères et sous-critères définis dans la norme ISO 19157 ne doivent pas être systématiquement utilisés dans tous les cas <i>La norme permet de s'appuyer sur les éléments existants, on prend ce qui est nécessaire</i> |
| <input type="checkbox"/> | La norme ISO 19157 est meilleure que la norme 19115 <i>Il n'y a pas de norme meilleur qu'une autre, chaque norme répond à une spécificité.</i> |
| <input checked="" type="checkbox"/> | Le contenu et la spécificité de la base de données permettent de savoir quels sont les critères les mieux adaptés pour en contrôler la qualité <i>Le contrôle de la qualité doit être en cohérent avec la contenu et la spécificité de la base de donnée</i> |
| <input checked="" type="checkbox"/> | Il est important de connaître les différents critères proposés pour assurer leur adéquation avec les différents types de contrôles à mettre en place <i>Les différents types de contrôles à mettre en place sont toujours liés aux critères qualités c'est qui permet à posteriori d'effectuer l'opération de contrôle.</i> |
| <input type="checkbox"/> | La mise en pratique des normes se fait uniquement par les experts <i>La mise en pratique des normes se fait pas tous, c'est pour cela que vous suivez cette formation</i> |

> **Solution n°3** (exercice p. 25)

| | |
|--------------------------------------|---|
| Le champ d'application de la mesure | C'est ce que l'on s'engage à mesurer |
| Le type de mesure effectuée | C'est la description de la formule utilisée |
| La procédure utilisée pour la mesure | C'est la méthode employée pour détecter les anomalies éventuelles |
| Le type de valeur et l'unité | C'est l'expression du résultat de la mesure de qualité |
| La date du contrôle | C'est la date à laquelle est effectuée le contrôle |

> **Solution n°4** (exercice p. 26)

| | |
|-------------------------------------|---|
| <input type="checkbox"/> | <p>La norme ISO 19157 est meilleure que la norme 19115 <i>Il n'y a pas de norme meilleure qu'une autre, chaque norme répond à une spécificité.</i></p> |
| <input checked="" type="checkbox"/> | <p>La norme ISO 19157 propose les différentes mesures qu'il est possible de réaliser par critère <i>La norme ISO 19157 propose les différentes mesures par critères :</i></p> <ul style="list-style-type: none"> • <i>Le champ d'application de la mesure</i> • <i>Le type de mesure effectuée</i> • <i>La procédure utilisée pour la mesure</i> • <i>Le type de valeur de l'unité</i> • <i>La date du contrôle qualité effectué sur l'indicateur</i> |
| <input checked="" type="checkbox"/> | <p>On choisit la ou les mesures les mieux adaptées en fonction du type de donnée à évaluer. Certains mesures sont très simples comme un simple comptage, une moyenne... D'autres sont plus complexes et nécessitent des mesures statistiques. <i>La type de mesure la mieux adaptée est la description de la formule présente dans le jeu de données cible et dans le jeu de contrôle</i></p> |
| <input type="checkbox"/> | <p>Il est important de connaître les différents critères proposés pour assurer leur adéquation avec les différents types de contrôle. On choisit la ou les mesures les mieux adaptées en fonction du type de donnée à évaluer. Pour chaque type de mesure, on applique des formules statistiques très compliquées <i>Pour chaque type de mesure on peut trouver un contrôle adapté sans qu'il soit très compliqué</i> <i>Cela peut-être :</i></p> <ul style="list-style-type: none"> • <i>un contrôle visuel par superposition à l'écran ;</i> • <i>un contrôle visuel par tirage papier (ou calque) ;</i> • <i>détection par requête spatiale (pour détecter les non-conformités de technologie par exemple) ;</i> • <i>par requête dans la base de données ;</i> • <i>...</i> <p><i>Le résultat obtenu</i></p> <ul style="list-style-type: none"> • <i>Il s'agit du résultat de la mesure spécifiant le type valeur et l'unité.</i> <i>Exemple : 95,7%</i> <i>Exemple : requête sql permettant de compter le nombre total d'objets surfaciques dans la commune X et le nombre d'objets affectés par l'erreur d'auto-intersection.</i> |
| <input type="checkbox"/> | <p>La mise en pratique de la norme 19157 se fait uniquement par les experts en statistique</p> |

> Solution n°5 (exercice p. 26)

| | |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | Indicateur d'erreur (booléen : la valeur est juste ou fausse) |
| <input checked="" type="checkbox"/> | Nombre d'erreurs : nombre total d'erreurs mesurées sur l'échantillon |
| <input type="checkbox"/> | Taille |
| <input type="checkbox"/> | Poids |
| <input checked="" type="checkbox"/> | Taux d'erreur : pourcentage d'erreurs relevées sur l'échantillon évalué |
| <input checked="" type="checkbox"/> | Matrice de co-variance |
| <input type="checkbox"/> | |