



Commission de validation des données pour l'information spatialisée



Comment qualifier la précision et les notions d'échelles dans les métadonnées de nos séries de données ?

version 3

Table des matières

Problématique et enjeu.....	1
La résolution spatiale.....	2
La précision de positionnement.....	3
La source des données.....	4
La généalogie de la ressource.....	4
Conseils de lecture.....	5

Problématique et enjeu

La description des données spatialisées est un exercice aujourd'hui indispensable pour qui veut échanger, diffuser ou ré-utiliser des données. Cette description se fait au moyen de métadonnées géographiques. Souvent qualifiées de « données sur les données », les métadonnées servent à la fois de fiche d'identité (identification) et de mode d'emploi (utilisation) des données géographiques. Des métadonnées bien renseignées sont une condition (nécessaire mais pas suffisante) pour qu'un utilisateur trouve, comprenne et sache utiliser les données dont il a besoin.

D'abord normalisées par l'ISO puis rendues obligatoires pour partie d'entre elles par la directive INSPIRE, certaines métadonnées géographiques demeurent pourtant difficiles à appréhender et a fortiori à renseigner. Quel administrateur de données ne s'est jamais interrogé sur comment remplir tel ou tel champ de son formulaire de métadonnées ? Si la création de nos métadonnées s'apparente trop souvent à de l'improvisation, c'est en partie à cause des normes et autres règlements qui donnent à lire des définitions laconiques ou subjectives, donc sujettes à interprétation. Dès lors, rien ne garantit qu'une même métadonnée soit toujours comprise et renseignée de la même façon. La COVADIS a fait le constat que les métadonnées d'un jeu de données peuvent même être incohérentes entre elles, en particulier les métadonnées de précision et d'échelles cartographiques d'utilisation.

Cette fiche présente les quatre métadonnées retenues par la COVADIS pour décrire la précision et les échelles d'utilisation d'un jeu de données numérique : la **résolution spatiale**, la **précision de positionnement**, la **source des données** et la **généalogie de la ressource**. La difficulté n'est pas de choisir – INSPIRE nous impose déjà des métadonnées – mais plutôt de rendre compréhensibles des notions qui viennent bousculer certaines habitudes.

La finalité de l'exercice est d'améliorer la qualité des métadonnées créées tandis que son enjeu est la confiance de l'utilisateur dans les données qu'il manipule. Cette fiche s'adresse autant aux personnes chargées de rédiger des métadonnées qu'aux utilisateurs qui les lisent. Certaines explications sont le fruit d'un travail commun au secrétariat de la COVADIS et au Pôle Géomatique du Ministère : elles proposent leur point de vue sur ces quatre métadonnées.

La résolution spatiale

Définition : La résolution spatiale se rapporte au niveau de détail de la série de données. Elle est exprimée comme un ensemble de valeurs de distance de résolution allant de zéro à plusieurs valeurs (utilisé pour des données maillées et des produits dérivés d'imagerie) ou exprimée en échelles équivalentes (habituellement utilisées pour les cartes ou les produits dérivés de cartes).

Qu'entend-on par résolution spatiale ?

La définition est assez vague et mérite quelques explications car le niveau de détail d'un jeu de données géographiques n'est pas du tout évident à déterminer. La norme ISO définit la résolution comme la métadonnée qui « fournit une idée sur la densité spatiale des données ». Si cette définition n'est guère plus précise que celle d'INSPIRE, elle est intéressante car elle relativise la portée de cette métadonnée. Une résolution spatiale n'est le résultat ni d'un calcul ni d'une mesure. Le producteur fournit une idée qui correspond à son appréciation de la densité spatiale de ses données. La valeur de cette métadonnée est une estimation subjective – donc discutable.

Comment renseigner la résolution spatiale ?

Dans le langage courant, les données spatiales sont souvent qualifiées de niveau communal, départemental, régional ou national. Si c'est pratique car compréhensible pour la communauté géomatique française, cela s'exporte assez mal. Le règlement INSPIRE exprime la résolution spatiale comme :

- une distance de résolution pour les données raster (distance sur le terrain du pas de la grille)
- le dénominateur de l'échelle équivalente pour des données vectorielles. L'échelle équivalente correspond à l'échelle la plus appropriée pour exploiter les données dans un SIG. Autrement dit, deux jeux de données de même résolution spatiale sont suffisamment comparables pour que leur croisement spatial soit pertinent.

Quelques résolutions de données vecteur sont données par le tableau suivant. Il s'agit d'ordres de grandeur estimés en fonction de la source des données utilisée pour leur géoréférencement.

Source de données utilisée en saisie		Résolution spatiale de la source des données	Ordre de grandeur de la résolution spatiale du jeu de données vecteur saisi	
RASTER	BD Parcellaire raster	1:500 → 1:625	0,05 m	2000
		1:1 000 → 1:1 250	0,1 m	
		1:2 000 → 1:2 500	0,2 m	
		1:4 000 → 1:5 000	0,5 m	
	BD ORTHO		0,2 – 0,3 – 0,4 ou 0,5 m	10 000
	SCAN 25		2,5 m	25 000
	SCAN Régional		25 m	250 000
SCAN 1000		100 m	1 000 000	
Document cartographique scanné		Dénominateur de l'échelle cartographique de la source	Dénominateur de l'échelle cartographique de la source	
VECTEUR	PCI / BD Parcellaire vecteur		Comprise entre le 500 et le 5 000	2000
	BD TOPO		10 000	10 000
	BD CARTO / BD CARTHAGE		50 000	50 000
	GEOFLA- communes		200 000	200 000
	ROUTE 500		200 000	200 000
	ROUTE 120		500 000	500 000

La précision de positionnement

Définition : Différence absolue entre les valeurs des coordonnées d'un jeu de données géographiques, et les valeurs (considérées comme ou étant) vraies.

Qu'entend-on par précision de positionnement ?

La précision de positionnement est un critère qualité quantitatif devant être le résultat d'une mesure. La précision de position se mesure en estimant les écarts entre la position des objets dans le jeu de données et la position vraie de leurs entités réelles – telle que mesurée par rapport au réseau géodésique. Cette mesure de nature statistique est en principe réalisée lors du contrôle qualité des données produites. Seulement, les contrôles qualité sont assez rares car difficiles à mettre en œuvre et coûteux.

Un producteur ne peut souvent en pratique que réaliser une estimation approchée de la précision géométrique de ses données.

Comment renseigner cette précision ?

A défaut de contrôle qualité, il n'y a pas de recette miracle. La métadonnée de précision doit être estimée par l'administrateur de données ou mieux par le producteur des données en fonction de leur connaissance des données. Cette estimation est un exercice subjectif qui tient compte :

- des caractéristiques du référentiel géographique ou de source de géoréférencement
- du mode opératoire, des traitements appliqués
- de la compétence et la rigueur de l'opérateur
- de la qualité graphique du document source

Le tableau ci-dessous fournit des exemples de précision de positionnement en fonction de la source du géoréférencement. Il s'agit d'ordres de grandeur (et non pas d'une mesure scientifique). Une bonne pratique consiste à afficher des valeurs plutôt moins-disantes, le but étant d'alerter un utilisateur pour qu'il évite de croiser des données géométriques de précisions trop différentes.

Source du géoréférencement	Précision de positionnement fournie dans les spécifications du référentiel	Ordre de grandeur estimé de la précision du jeu de données
BD Parcellaire	0,5m < emq < 5m (suivant la typologie de la feuille cadastrale source)	5 m
BD ORTHO	emq variable suivant les dalles	5 m
BD TOPO	En planimétrie, selon la source des données : <ul style="list-style-type: none"> • Levé GPS de précision : emq = 0,5 m • Photogrammétrie, plan ou fichier métrique : emq = 1,5 m • Photogrammétrie des PVA longue focale, levé GPS dynamique, levé terrain : emq = 2,5 m • Orthophotographie, plan ou fichier non métrique, BD TOPO® version antérieure : emq = 5 m • Carte 1/25000 : emq = 10 m • BD CARTO® : emq = 30 m 	5 m
MNT BD TOPO	Précision planimétrique : 1,5 m / Précision altimétrique : 2, 50 m à 4,00 m (selon la nature du relief)	1,5 m (planimétrie) 4m (altimétrie)
SCAN25	emq = 10 m (emq = 5 m pour les éléments les plus fins, dépourvus de surcharges cartographiques, n'ayant subi ni généralisation, ni décalage)	10 m
BD CARTO BD Carthage	15 m < emq < 50 m	50 m
GéoFLA	200 m < emq < 500 m	500 m
SCAN Régional	emq = 50 m mais il faut tenir compte du processus de généralisation qui introduit des écarts qui ne sont pas mesurés.	100 m
ROUTE 500	emq = 50 m pour les nœuds routiers emq = 100 m pour les nœuds des réseaux hydrographique et ferré, les sommets du graphe des limites départementales	100 m

ROUTE 120	emq = 50 m pour les nœuds routiers emq = 100 m pour les nœuds des réseaux hydrographiques et ferré, les sommets du graphe des limites départementales, les chefs-lieux de communes et les aérogares	100 m
-----------	--	-------

Emq : écart moyen quadratique

La précision d'une donnée ne correspond pas forcément à celle du référentiel utilisé même si elle est généralement étroitement liée. Elle sera au mieux celle des objets du référentiel sur lesquels l'opérateur s'est appuyé lors de la saisie.

La source des données

Définition : Nom et échelle de la carte ou du référentiel géographique utilisé comme source des données

Qu'entend-on par source des données ?

Les sources des données correspondent aux fichiers, bases de données et documents qui ont fourni l'information contenue dans un jeu de données. Les données géographiques ont la particularité de souvent associer à des données de géoréférencement (carte, référentiel géographique) des données applicatives ou thématiques, qui proviennent de sources différentes.

Comment renseigner cette métadonnée ?

Une bonne pratique est de distinguer les sources des données qui ont servi au géoréférencement de celles qui ont fourni l'information thématique.

Une source des données géoréférencée est décrite par son nom, sa résolution spatiale et sa date de mise à jour. Il convient de systématiquement renseigner cette source des données.

Une source des données thématique est décrite par son nom, sa date et, le cas échéant, son auteur ou son gestionnaire.

La généalogie de la ressource

Définition : La généalogie fait état de l'historique du traitement et/ou de la qualité générale de la série de données géographiques. Le cas échéant, elle peut inclure une information indiquant si la série de données a été validée ou soumise à un contrôle de qualité, s'il s'agit de la version officielle (dans le cas où il existe plusieurs versions) et si elle a une valeur légale.

Qu'entend-on par généalogie ?

Selon la norme ISO 19115, la généalogie de la ressource décrit l'historique d'un jeu de données et, s'il est connu, le cycle de vie de celui-ci, depuis l'acquisition et la saisie de l'information jusqu'à sa compilation avec d'autres jeux et les variantes de sa forme actuelle.

Comment renseigner cette métadonnée ?

Il s'agit d'apporter une description littérale et concise soit de l'histoire du jeu de données, soit des moyens, procédures ou traitements informatiques mis en œuvre au moment de l'acquisition du jeu de données.

Par exemple, la généalogie peut consigner l'échelle de saisie si cette information est importante pour l'utilisation du jeu de données.

Les métadonnées source des données et généalogie de la ressource sont complémentaires et donnent une indication de la qualité globale du jeu de données.

Conseils de lecture

Alors que le partage des données se généralise sous l'impulsion du législateur, que les nouvelles infrastructures de données offrent de plus en plus de données à la demande, un internaute peut aujourd'hui facilement télécharger, covisualiser, superposer et combiner des données géographiques parfois incohérentes, de qualité hétérogène. Des métadonnées simples et fiables sont plus que jamais nécessaires pour encadrer ces échanges de données et prévenir les mauvaises utilisations. Quelques conseils de lecture s'imposent pour faire évoluer nos habitudes et savoir bien lire ces quatre métadonnées.

Changer son point de vue

Tout d'abord, aucune des quatre métadonnées retenues n'utilise plus le mot échelle. Le terme *résolution* semble plus approprié pour décrire des données géographiques que le terme *échelle* qui se rapporte plus à la représentation (sur un écran, sur une carte) des données. Même si la notion d'échelle tient encore une grande place dans nos habitudes de travail, le nouveau vocabulaire marque un changement de point de vue. La précédente norme européenne ENV 12657 – utilisée par REPORTS – décrivait les données avec une approche plus cartographique : il fallait renseigner un intervalle de visualisation déterminé par une échelle minimale et une échelle maximale. Les données géographiques sont numériques et nous obligent à changer de vocabulaire et d'habitude. Les métadonnées retenues par la directive INSPIRE se veulent désormais adaptées aux outils géomatiques et autres SIG dans lesquels une échelle cartographique n'est qu'un paramètre d'affichage et d'édition pour représenter un ou plusieurs jeux de données géographiques.

Estimer puis relativiser

Cela a-t-il du sens de dire que la BD Parcellaire est une base de données cadastrales à l'échelle du 1 :2000 ? Un géomètre-expert peut considérer cette affirmation caricaturale quand l'échelle d'un levé peut varier entre le 1 :5000 et le 1 :500 ; tandis que l'utilisateur d'un outil SIG interprète cette information comme une des échelles possibles de visualisation.

Est-il juste de dire que des données saisies sur un SCAN25® ont une précision de positionnement de 5 mètres ? Le SCAN25® est un produit cartographique qui n'est pas un référentiel géographique. 5 mètres est la meilleure précision de positionnement d'un objet numérisé dans une zone peu dense sur les figurés de la carte les plus fins, dépourvus de surcharges cartographiques, n'ayant subi ni généralisation, ni décalage, rendus nécessaires par l'édition papier. Par prudence, la précision de positionnement des données sera plutôt estimée à 10 mètres.

Des données saisies en s'appuyant sur le réseau routier de la BDTOPO® ont-elles une précision de positionnement égale à un mètre ? La précision d'un objet géométrique dans le jeu de données obtenu est égale à celle du tronçon routier sur lequel l'opérateur s'est appuyé – à condition de n'avoir pas fait d'erreur de saisie grossière. Or les spécifications actuelles de la BDTOPO® précisent que les objets tronçon de route ont une précision de positionnement variable suivant la source géométrique utilisée (cf. tableau page 3). Donc la précision de positionnement globale du jeu de données produit sera estimée à 5 mètres par prudence, sauf à entreprendre un contrôle qualité rigoureux. L'utilisateur saura qu'il ne peut envisager croiser ces données avec un autre jeu de 100 mètres de précision sans prendre de précautions. La métadonnée de précision joue un rôle d'alerte et de garde-fou.

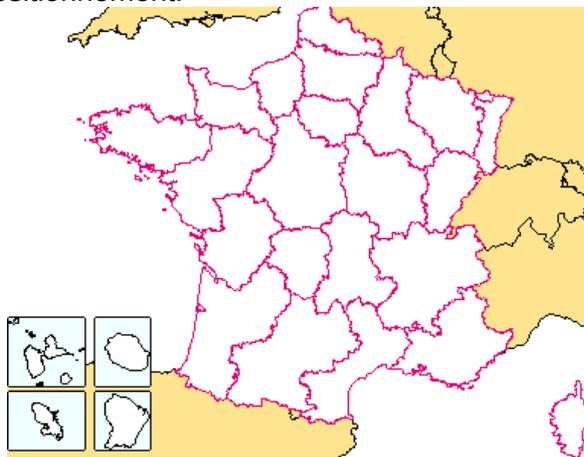
Ces quelques cas montrent que l'appréciation des métadonnées de résolution et de précision est souvent un exercice à la fois subjectif et complexe. Chaque producteur a tendance à décrire ses données en fonction des usages qu'il en fait. Comment un utilisateur va-t-il alors être capable d'évaluer si un jeu de données répond à son besoin ? D'un côté, le producteur estime l'ordre de grandeur de la résolution et de la précision de son jeu de données. De l'autre côté, un utilisateur averti doit relativiser les valeurs de métadonnées dont il dispose. Il lui faut considérer les métadonnées non plus comme des descripteurs absolument vrais, scientifiquement mesurés ou calculés, mais comme des indicateurs de qualité qui n'ont d'intérêt que s'ils sont comparés d'un jeu de données à l'autre. De plus, ne renseigner que la résolution spatiale d'un jeu de données ne suffit pas à ce qu'un utilisateur puisse en apprécier les utilisations potentielles. C'est pour cette raison qu'il faut systématiquement compléter les quatre métadonnées « précision, résolution, source, généalogie » de façon cohérente.

Des métadonnées complémentaires

La source des données, la généalogie et la précision de positionnement sont des métadonnées complémentaires – surtout quand on décrit des bases de données géographiques. Une base de données est souvent le produit complexe d'un processus qui a traité des informations de nature et de sources variées. Sans mesure de la qualité, il est difficile de se faire une idée de la précision ou la résolution de cet ensemble de données composites. Décrire finement la qualité géométrique d'une base de données passe par la gestion de métadonnées au niveau de l'objet – à l'instar du champ 'source_geometrie' utilisé dans chaque fichier de la BDTOPPO®. Mais une telle granularité est assez rare dans les données thématiques tandis qu'une unique valeur de précision de positionnement pour qualifier toute une base de données peut sembler suspecte. La description de la généalogie et des sources des données vont alors servir à bien apprécier la précision globale de positionnement et la résolution spatiale de la base de données. Une lecture combinée de ces métadonnées fournit en général une bonne indication de la qualité globale du jeu de données, donc de son utilisation potentielle

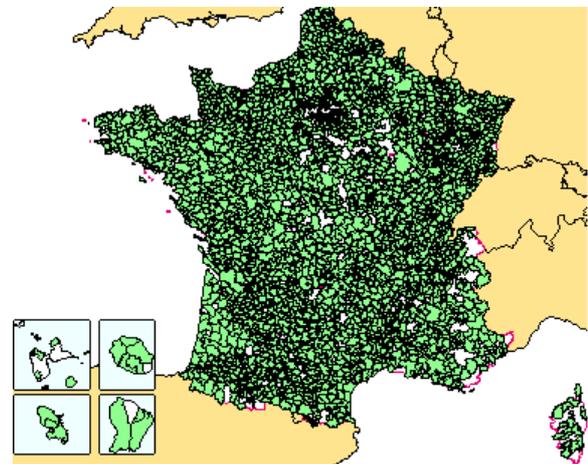
Précision versus résolution : ne pas confondre !

Résolution spatiale ne veut pas dire précision géométrique ! Ce sont deux caractéristiques différentes peuvent ne pas être corrélées. L'exemple ci-dessous réalisé avec les données de la base GEOFLA de l'IGN montre que la résolution spatiale ne rend pas compte de la précision. Il est possible d'augmenter la densité spatiale des informations sans en améliorer la précision de positionnement.



Limites administratives régionales

Source : GEOFLA®, protocole MEEDDAT - MAP - IGN du 24 juillet 2007



Limites intercommunales