

Les rapports d'étude

La qualité des données géographiques

État des lieux pour un débat



Présent
pour
l'avenir



|

La qualité des données géographiques

État des lieux pour un débat

décembre 2010



centre d'Études sur les réseaux,
les transports, l'urbanisme
et les constructions publiques
9 rue Juliette Récamier
69456 Lyon Cedex 06
téléphone : 04 72 74 58 00
télécopie : 04 72 74 59 00
www.certu.fr

Avis aux lecteurs

La collection Rapports d'étude du Certu se compose de publications proposant des informations inédites, analysant et explorant de nouveaux champs d'investigation. Cependant l'évolution des idées est susceptible de remettre en cause le contenu de ces rapports.

Le Certu publie aussi les collections :

Dossiers : Ouvrages faisant le point sur un sujet précis assez limité, correspondant soit à une technique nouvelle, soit à un problème nouveau non traité dans la littérature courante. Le sujet de l'ouvrage s'adresse plutôt aux professionnels confirmés. Ils pourront y trouver des repères qui les aideront dans leur démarche. Mais le contenu présenté ne doit pas être considéré comme une recommandation à appliquer sans discernement, et des solutions différentes pourront être adoptées selon les circonstances.

Références : Cette collection comporte les guides techniques, les ouvrages méthodologiques et les autres ouvrages qui, sur un champ donné, présentent de manière pédagogique ce que le professionnel doit savoir. Le Certu a suivi une démarche de validation du contenu et atteste que celui-ci reflète l'état de l'art. Il recommande au professionnel de ne pas s'écarter des solutions préconisées dans le document sans avoir pris l'avis d'experts reconnus.

Débats : Publications recueillant des contributions d'experts d'origines diverses, autour d'un thème spécifique. Les contributions présentées n'engagent que leurs auteurs.

Catalogue des publications disponible sur : <http://www.certu.fr/catalogue>

Organisme commanditaire : Certu			
Titre : La qualité des données géographiques			
Sous-titre : État des lieux pour un débat	Date d'achèvement : 29/10/10	Langue : Français	
Organisme auteur : Certu	Rédacteurs ou coordonnateurs : Gilles Troispoux	Relecteur assurance qualité : Gilbert Nicolle, Jean-Loup Delaveau, Élisabeth Rinié (CGEDD)	
<p>Résumé</p> <p>Le développement considérable de l'information géographique, que ce soit dans un contexte professionnel ou pour une utilisation « grand public », nous oblige à reconsidérer de plus en plus sérieusement le problème de la qualité des données dont l'impact a une influence directe sur la fiabilité des analyses spatiales produites et des décisions qui en découlent.</p> <p>Ce rapport 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.</p> <p>La notion de responsabilité et les risques encourus de mauvaise utilisation de données mal renseignées sont incompatibles avec le développement de l'information géographique tant au niveau professionnel que du grand public.</p> <p>L'objectif principal de ce rapport 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.</p>			
Observations :			
A mettre sur le portail de l'environnement : oui <input checked="" type="checkbox"/> non <input type="checkbox"/>			
Thème et sous-thème :			
Information géographique : <ul style="list-style-type: none"> - données, méthodes et standardisation (web Certu) ; - données de références et méthodes pour l'utilisation de l'IG (Procert). 			
Ouvrages sur le même thème : La qualité des données géographiques (CD ROM 2006)			
Mots clés : données géographiques, qualité (des données géographiques), métadonnées, catalogage, ISO, INSPIRE, COVADIS, Référentiel géographique, progiciels géomatiques, SIG, formation continue, formation initiale		Diffusion : Web	Web : oui
Nombre de pages : 29	Tirage : non	Confidentialité : non	Bibliographie : non

Remerciements

Ce document est né à la suite de nombreux échanges menés avec des personnes convaincues de l'importance de la qualité des données géographiques. Afin de n'oublier personne, nous ne les citerons pas nommément, ils se reconnaîtront et nous les en remercions vivement. Les organismes concernés sont les suivants : COGIT, IGN, réseau canadien GEOIDE, CRIGE PACA, SI 17, GAIAGO, IETI consultants, AFIGEO, etc.

Sommaire

1. Introduction	6
2. De la théorie à la pratique	7
2.1 La qualité interne	7
2.2 La qualité externe	7
3. Dysfonctionnements perçus autour de la qualité des données	9
3.1 La qualité des référentiels géographiques	9
3.2 La complexité des outils et des règles	11
3.3 Des compétences et une sensibilité des utilisateurs à développer	16
3.4 Un monde en perpétuelle évolution	18
3.5 Les métadonnées perçues comme un échec partiel	18
3.6 Une terminologie à redéfinir	21
3.7 La qualité : à la frontière entre plusieurs cultures	22
4. Vers une nouvelle doctrine de la qualité	24
4.1 Le tableau de bord et les indicateurs de qualité	24
4.2 Le « geotagging » ou Post-It® géographique	24
4.3 Les réseaux sociaux professionnels	25
4.4 Les spécifications initiales et l'expression des besoins	25
5. Conclusion	26
6. Bibliographie	27
7. Table des matières	28

1. Introduction

Le développement considérable de l'information géographique, que ce soit dans un contexte professionnel ou pour une utilisation « grand public », nous oblige à reconsidérer de plus en plus sérieusement le problème de la qualité des données dont l'impact a une influence directe sur la fiabilité des analyses spatiales produites et des décisions qui en découlent. Une autre question qui est soulevée résulte d'un besoin impératif qui oblige producteurs et distributeurs à informer les utilisateurs de la qualité des données spatiales d'une manière significative et compréhensible (comme n'importe quel autre produit ou service sur le marché).

Les outils actuels qui permettent de mesurer, qualifier et communiquer la qualité des données semblent inadaptés aux besoins, aux enjeux et à la diversité des acteurs qui composent la communauté géomatique.

A l'issue de quelques entretiens, rencontres et avis divers, ce document se propose de faire un constat des différents problèmes mis en évidence. Il n'a aucune prétention ni de rigueur scientifique ni d'exhaustivité mais une simple ambition d'initier le débat au sein de la communauté géomatique. Producteurs, distributeurs, utilisateurs, chercheurs ont très certainement de nombreux avis à émettre pour l'alimenter et le faire vivre afin de déboucher sur des actions pragmatiques et opérationnelles.

2. De la théorie à la pratique

Un rapide rappel sur la théorie s'avère nécessaire pour fixer le vocabulaire, rappeler le contexte normatif et souligner que la réalité opérationnelle de certains producteurs de données et de la communauté des utilisateurs est très éloignée des définitions académiques de la qualité des données.

Le monde de la recherche propose deux définitions pour la qualité des données géographiques.

2.1 La qualité interne

On peut traduire la qualité interne de la sorte : « comment puis-je mesurer la qualité de mes données et comment le faire savoir ? ». Elle se mesure au travers de critères (précision géométrique, exhaustivité, précision sémantique, cohérence logique, actualité, etc.) définis dans la norme ISO 19113 à l'aide de méthodes explicitées dans la norme ISO 19114 et de mesures spécifiques dans la norme ISO 19138.

Pour communiquer cette évaluation de la qualité, on utilise alors les métadonnées de la norme ISO 19115 ou on produit un rapport spécifique sur la qualité qui demeure, dans la plupart des cas, un outil d'expert. On rappelle que toutes les déclarations sur la qualité des données sont optionnelles au sein de la norme ISO 19115, ce qui n'incite pas vraiment les producteurs à diffuser les informations sur la qualité. Une nouvelle norme (ISO 19157) est en cours de rédaction et devrait remplacer prochainement les normes ISO 19113, 19114 et 19138.

La directive INSPIRE doit revoir à la hausse certaines obligations en terme de diffusion d'informations liées à la qualité.

Ces outils demeurent « la référence » pour évaluer la qualité d'un lot de données mais s'avèrent limités car ils nécessitent de connaître les **spécifications initiales** de la base de données que l'on désire contrôler et s'appuient sur des techniques professionnelles. Ces normes sont le fruit d'un travail commun mené par les organismes nationaux de cartographie au sein des institutions de normalisation. Si elles répondent en grande partie aux besoins de ces organismes qui produisent les données de référence, elles semblent plutôt mal adaptées aux besoins des utilisateurs et encore moins à l'évolution « grand public » que prend la géomatique aujourd'hui autour de l'internet et des productions collaboratives.

2.2 La qualité externe

Elle se définit plutôt comme une adéquation aux besoins que l'on peut résumer ainsi : « quels sont les besoins des utilisateurs en termes de données et d'informations de qualité et comment puis-je les leur offrir pour qu'ils évitent les utilisations abusives de ces données ? ». Cette approche que le monde anglophone résume par l'expression « fitness for use » et que l'on peut traduire comme une « aptitude à l'emploi » crée un lien plus direct entre les ensembles de données et leurs utilisations possibles, entre les préoccupations des producteurs et les attentes de la plupart des utilisateurs. Mais la qualité externe reste plus difficile à définir eu égard aux besoins qui demeurent multiples et variés.

Cette approche relève encore du domaine de la recherche et aucune méthode ou outil n'est véritablement parvenu jusqu'aux utilisateurs.

Pour résumer, la qualité interne exprime plus la qualité intrinsèque d'un lot de données exprimée au travers de règles précises alors que la qualité externe exprime l'adéquation à un besoin, la capacité à répondre à un usage particulier. On pourrait parler de qualité technique absolue et de qualité d'usage relative.

Ainsi, les pratiques liées à la qualité et à sa communication sont encore très embryonnaires et relèvent exclusivement de la « qualité interne ». Elles se limitent aux quelques champs des métadonnées que les producteurs ont bien voulu renseigner mais se heurtent souvent à des problèmes de terminologie, de coût et de compétences. Nous reviendrons plus en détail sur ces considérations dans le chapitre suivant.

3. Dysfonctionnements perçus autour de la qualité des données

La communauté géomatique maîtrise plutôt mal ces aspects liés à la qualité des données géographiques. Les raisons sont multiples. Il paraît opportun d'inventorier les causes les plus reconnues et tenter de proposer quelques embryons de solutions.

3.1 La qualité des référentiels géographiques

3.1.1 Le constat

La qualité des données métier produites par les services utilisateurs est très dépendante de la qualité des données géographiques de référence utilisées. Elles sont généralement produites par numérisation soit sur un support raster (orthophotographie, carte, plan, etc.), soit en dupliquant tout ou partie des données vectorielles. On mesure alors l'importance des référentiels partagés par la communauté géomatique.

La notion de référentiel demeure un concept simple qui a été longuement débattu et décrit au sein d'une commission éponyme du CNIG et de ses groupes de travail associés. Rappelons quelques règles qui les caractérisent : les référentiels géographiques sont clairement identifiés, ils sont produits par une structure publique reconnue, ce qui leur confère un niveau de confiance élevé des utilisateurs. Leur couverture est exhaustive sur le territoire. Leur contenu doit être limité afin de minimiser les interprétations possibles, mais doit pouvoir offrir, la géométrie et le géoréférencement, accompagnés d'une sémantique épurée et plutôt minimaliste qui correspond plus à l'intersection des besoins qu'à leur union. Une sémantique riche et complexe relève plutôt du référentiel thématique dont l'objectif est de fournir une information sémantique. La qualité de ces référentiels doit être connue et diffusée par leur producteur selon les normes en vigueur.

On comprend alors combien la notion de qualité inhérente à ces référentiels géographiques et thématiques est importante. Une donnée métier produite hérite en grande partie de la géométrie, du géoréférencement et des informations sémantiques contenues dans les référentiels sur lesquels elle s'appuie. Plus les données de référence sont exemptes de défauts, plus la probabilité d'avoir des données métier de bonne qualité est grande. Le monde de la recherche travaille depuis longtemps sur la propagation des erreurs... Dans un monde parfait, on pourrait alors rêver de la propagation de la qualité !

Les référentiels géographiques les plus utilisés en France, sont les données de l'IGN et le plan cadastral.

Le rapport Lengagne a contraint l'IGN à produire les quatre composantes du référentiel à grande échelle (RGE) dans un temps record, ce qui ne s'est pas fait sans compromis sur le plan de la qualité. De plus, on constate actuellement que les métadonnées décrivant la qualité et devant être diffusées avec les données sont quasiment absentes à cause, certainement, du délai très court pour les produire et de la non diffusion actuelle, par l'IGN, des conclusions de ses contrôles qualité. Les seules informations de qualité qui accompagnent pour le moment le RGE sont contenues dans ses spécifications. Or ces spécifications expriment l'objectif à atteindre et non pas dans quelles proportions il a pu être atteint. Autrement dit, tant

que des contrôles n'auront pas été réalisés et leurs résultats communiqués, l'utilisateur ne sera pas en mesure de se faire une idée précise de la qualité des référentiels qu'il manipule.

Pour le plan cadastral, on sait combien sa qualité est variable sur tout le territoire. Cette réalité est très regrettable car il demeure très largement utilisé au sein des collectivités territoriales, qui constituent un acteur très important de la communauté géomatique française. En plus de sa qualité hétérogène, le plan cadastral offre deux géométries différentes, l'une au sein du PCI (plan cadastral informatisé) et l'autre, dans la BD PARCELLAIRE® de l'IGN. Le PCI est utilisé essentiellement par les collectivités territoriales, alors que la BD PARCELLAIRE® de l'IGN l'est par les services de l'État. Cette antinomie ne va pas vraiment dans le sens de l'amélioration de la qualité des données métiers produites à partir de ce référentiel à géométrie variable et perturbe considérablement les échanges de données.

3.1.2 Les améliorations possibles

Les référentiels géographiques

Produire de telles données, avec certaines exigences coûte très cher si l'on veut obtenir une qualité définie et homogène sur tout le territoire avec des mises à jour régulières. Il faut donc bien cerner le but à atteindre et s'y limiter strictement.

L'objectif d'un référentiel géographique est de fournir de la géométrie et du géoréférencement, le tout accompagné d'un minimum de sémantique, de métadonnées et d'information sur la qualité. En ce qui concerne la sémantique, on ne peut effectivement pas saisir une ligne ou une surface sans un minimum de description, il faut bien pouvoir différencier une route d'une rivière, un bâtiment d'une surface boisée ou hydrographique. En revanche, la limite d'information sémantique à offrir est difficile à définir. La première version du RGE® est maintenant terminée. On sait que l'IGN est sollicité de toutes parts pour saisir de nombreuses données supplémentaires dans la BD TOPO®. Il faut rappeler qu'un référentiel géographique ne doit pas contenir l'union des besoins, mais leur intersection. Il serait peut-être opportun voire nécessaire que producteurs et utilisateurs se réunissent à nouveau et fixent ensemble le contenu du RGE® version 2 en séparant les informations qui relèvent d'un référentiel géographique de celles qui appartiennent à un référentiel métier ou thématique.

La commission des référentiels du CNIG serait un lieu idéal pour de tels débats sur le contenu des référentiels géographiques.

En attendant la mise à disposition de référentiels géographiques répondant à la grande majorité des besoins, gageons que les producteurs auront la volonté d'améliorer en continu leurs référentiels, mais surtout de diffuser en parallèle les éléments de qualité qui sont indispensables pour une meilleure prise en main et une utilisation plus rationnelle.

A court terme, deux décisions semblent s'imposer :

diffuser les métadonnées de qualité des référentiels qui sont indispensables aux utilisateurs pour mieux les maîtriser ;

faire en sorte qu'une convergence entre le PCI et la BD PARCELLAIRE® intervienne le plus rapidement possible.

Sur un plus long terme, il faudra

améliorer progressivement la qualité géométrique du plan cadastral.

Les référentiels thématiques

Alors que la notion de référentiel géographique impacte essentiellement la géométrie, les référentiels thématiques ou métier s'intéressent plutôt à la sémantique. Les producteurs de référentiels géographiques se comptent sur les doigts d'une main. En revanche, de nombreux acteurs interviennent sur des données thématiques dans des domaines très nombreux, très variés et très imbriqués.

De nombreuses bases de données thématiques ont déjà vu le jour sur des emprises locales mais il existe trop peu d'initiatives nationales pour constituer de véritables référentiels construits sur un modèle commun et des spécifications nationales. La directive INSPIRE pour les données liées à l'environnement et les travaux de la COVADIS (commission de validation des données pour l'information spatialisée) pour les données utilisées par les ministères de l'écologie et de l'agriculture sont deux démarches qui devraient améliorer l'utilisation, la production et les échanges de données.

On soulignera également l'initiative du CNIG qui a produit un premier rapport provisoire, issu de son groupe de travail « communautés thématiques – pôles métiers », destiné à accompagner les velléités naissantes d'échange et de partage de données métier.

De telles démarches sont à encourager et à poursuivre en sensibilisant fortement la communauté géomatique.

Sensibiliser l'ensemble des acteurs responsables d'une thématique donnée à spécifier, modéliser et produire des données de qualité.

Mettre en place un groupe de travail sur le sujet disposant de moyens et de ressources.

3.2 La complexité des outils et des règles

3.2.1 Le constat

A ce jour, nous sommes en mesure d'évaluer la qualité interne d'un jeu de données, mais à condition que ces données aient été préalablement spécifiées et en utilisant des outils complexes définis dans les normes internationales.

3.2.1.1 La réalité normative

Aux référentiels qui ne remplissent pas toujours leur rôle, il faut ajouter la complexité des outils normatifs difficiles à mettre en œuvre tant par les producteurs que par les utilisateurs. Trois normes ISO permettent de décrire les éléments de qualité et de les mesurer : les normes ISO 19113, 19114 et 19138. Elles seront très prochainement remplacées par une unique norme : la norme ISO 19157.

La **norme ISO 19113** décrit les différents critères de qualité d'une base de données géographiques. Ces différents critères sont les suivants :

Critères quantitatifs

- **Précision géométrique** : elle décrit la précision de positionnement planimétrique et altimétrique des données.

- **Exhaustivité** : elle décrit la présence ou l'absence d'objets, de leurs attributs ou de leurs relations dans le jeu de données.
- **Précision sémantique** : elle décrit la justesse des valeurs des éléments du jeu de données au niveau des objets, des attributs ou des relations.
- **Cohérence logique** : elle décrit le degré de compatibilité avec les règles logiques de structure de données, d'attributs et de relations (la structure peut être conceptuelle, logique ou physique).

Critères qualitatifs

- **Actualité** : elle introduit une référence temporelle permettant de savoir si les données sont à jour.
- **Généalogie** : la généalogie 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. Ce critère devrait ne plus être présent dans la nouvelle norme ISO 19157. En revanche, il reste bien présent dans la norme ISO 19115 au niveau des métadonnées.

Critère spécifique

- **Qualité spécifique** : la qualité spécifique permet à l'utilisateur de définir ses propres critères de qualité si les critères officiels définis par la norme ne répondent pas à ses attentes. C'est en quelque sorte un "critère personnalisé". Ce critère devrait ne plus être présent dans la nouvelle norme ISO 19157.

La **norme ISO 19114** fournit un cadre de procédures, de méthodes, de déterminations et d'évaluations de la qualité compatibles avec les principes et critères définis dans la norme ISO 19113. Elle établit également un cadre pour évaluer et rendre compte des résultats, soit par l'intermédiaire de métadonnées seulement, soit dans un rapport spécifique d'évaluation de la qualité.

La **norme ISO 19138** décrit un ensemble de mesures de la qualité des données. Elles peuvent être utilisées pour l'estimation des critères identifiés dans la norme ISO 19113. De nombreuses mesures sont définies pour chaque critère de la qualité des données. Le choix de les utiliser ou non dépend du type de la donnée à évaluer et de son usage.

Ces critères et procédures permettent la description qualitative d'un jeu de données à partir du moment où l'on connaît les objectifs finaux de sa production qui doivent être consignés dans un cahier des charges ou dans des spécifications. Une fois ces spécifications rédigées et les données produites, la qualité du jeu de données se mesure en comparant les données saisies aux ambitions initialement prévues. La qualité est alors consignée dans un rapport spécifique ou à l'aide des métadonnées définies dans la **norme ISO 19115**.

Ce processus est parfaitement adapté à un contexte professionnel de production de données, mais reste, cependant, particulièrement lourd à mettre en œuvre car les normes ISO 19114 et ISO 19138 s'appuient sur des démarches statistiques et topographiques qui nécessitent un niveau d'expertise et un investissement élevés.

3.2.1.2 Manque de spécifications initiales

Sans la connaissance des spécifications initiales, il n'est pas possible, de mesurer la qualité interne d'une base de données à l'aide des outils normatifs décrits ci-dessus.

Comment apprécier la qualité de telles données non spécifiées ? Il faudrait pouvoir les comparer à des données pouvant servir de référence dans la mesure où leur qualité est parfaitement connue et a priori considérée comme meilleure (comparaison de la géométrie avec celle de référentiels géographiques reconnus, et de la sémantique avec des bases de données métier également reconnues). Cet exercice réalisé au Certu autour d'une base de données nationale sur les transports publics a très vite montré ses limites tant pour la qualité géométrique que sémantique des données. Dans les deux cas, il fut difficile de disposer des données de référence reconnues et fiables pour effectuer une comparaison. De plus, les méthodes imaginées sont devenues très rapidement complexes en dehors de tout cadre précis de définition et de production.

Ainsi, que l'on dispose ou non de spécifications initiales, la démarche reste complexe dans tous les cas pour évaluer la qualité d'un jeu de données. En revanche, les outils et méthodes existent si l'on a pris la précaution de spécifier ses données avant d'entreprendre une production de données. Cela milite fortement pour motiver tout producteur potentiel de données à les spécifier à l'aide d'une méthode reconnue.

La plupart des utilisateurs qui désirent produire des données métier, n'adoptent pas toujours cette démarche méthodologique. Dans ces conditions, les données produites répondent uniquement à leurs besoins immédiats, mais deviennent rapidement caduques pour eux-mêmes par perte de mémoire inévitable et quasiment inutilisables par d'autres services.

Cette constatation est aujourd'hui une évidence au sein des différentes dynamiques régionales (CRIGE PACA, Syndicat informatique 17, etc.) dont l'objectif était, dans un premier temps d'acquérir des référentiels communs en partageant les coûts et dans un second temps, de pouvoir échanger des données métier, véritable matière première des thématiciens. Cette seconde ambition n'est pas envisageable tant qu'on sera dans l'impossibilité de mesurer et de communiquer la qualité des données produites à cause de l'absence de spécifications.

La norme qui définit les spécifications de contenu informationnel est la norme ISO 19131.

3.2.2 Les améliorations possibles

3.2.2.1 Proposer un outillage plus simple

Produites par des experts pour des experts, les normes sont plutôt adaptées à l'évaluation des données produites dans un contexte professionnel. Il serait inopportun de les remettre en cause. En revanche, elles demeurent très peu utilisées, même par les organismes responsables de la production des données de référence, car elles nécessitent la mise en place de méthodes lourdes et coûteuses. La lecture et l'interprétation des résultats des contrôles qualité effectués selon ces normes par le monde des utilisateurs, nécessitent également un bagage technique certain. Tout processus de communication entre producteurs de signaux et récepteurs - individus ou machines - nécessite des connaissances partagées entre producteur et récepteur.

- Ces normes sont-elles adaptées à la communauté géomatique dans son ensemble et dans sa diversité ?

Si la réponse est positive, il faut alors se donner tous les moyens pour que ces outils et ce langage soient pleinement utilisés et surtout compris par tous. Un tel statu quo impose d'importants efforts pour les producteurs, qu'ils soient professionnels ou pas, nécessite de définir des obligations, de mettre en place de nombreuses mesures d'accompagnement sans pour autant être une garantie de réussite à l'image des problèmes rencontrés actuellement.

Mise en place de mesures d'accompagnement, sensibilisation, formation, assistance, recours à des experts, etc.

Le Certu a produit à cet effet un CD-ROM qui présente les deux normes ISO 19113 et ISO 19114 sous une forme simplifiée et vulgarisée. On peut acquérir ce CD à l'adresse suivante :

http://www.certu.fr/catalogue/product_info.php?products_id=1078&language=fr

Avec de telles mesures, on peut espérer voir le niveau de professionnalisme s'accroître, et par là même, le niveau de qualité des données.

Dans le cas d'une réponse négative, d'autres questions se posent :

- Faut-il simplifier ces normes ?
- Faut-il définir des outils adaptés à chaque type d'intervenants ?
- Faut-il être plus péremptoire et imposer certains contrôles et certaines métadonnées ?
- Faut-il mettre en place une assistance aux utilisateurs ?
- Peut-on conserver ces outils pour le monde professionnel et envisager une version simplifiée, une sorte de filtre, pour le monde utilisateur et/ou grand public ?

Toutes ces interrogations nécessitent un véritable débat au sein de la communauté géomatique.

3.2.2.2 Apprendre à spécifier et à contrôler les données

Comme nous l'avons vu, une vraie démarche professionnelle consiste, avant toute production de données, à rédiger des spécifications tant pour les données que pour les techniques de saisie. Cette méthode est utilisée par les producteurs institutionnels de données référentielles, c'est à dire par le monde professionnel, et devrait être pratiquée systématiquement pour toute autre saisie de données thématiques ou métier.

En dehors d'une telle démarche, il devient très difficile et très coûteux de pouvoir vérifier la qualité d'un lot de données.

La première question que l'on peut alors se poser est donc :

- Pourquoi ne rédige-t-on pas systématiquement les spécifications d'un lot de données avant de le produire ?

La réponse est certainement multiple : par manque de temps, par manque de compétence et d'expérience, par méconnaissance des pratiques, etc.

Dans ce cas, la meilleure solution consiste à informer, sensibiliser, former et assister. On peut donc imaginer les réalisations suivantes :

guide pour la rédaction de spécifications issu de la norme ISO 19131. Des travaux importants ont déjà été menés dans le cadre d'INSPIRE ;

mise en place de modules de formations ;

sensibilisation des acteurs par des organismes institutionnels ou associatifs comme le CNIG ou l'AFIGEO, etc.

assistance technique.

La mise en place de la directive INSPIRE devrait largement contribuer à professionnaliser les différents acteurs et notamment ceux qui devront diffuser leurs données. Cela n'exclut pas que des mesures d'accompagnement sont nécessaires pour améliorer la production des données thématiques.

3.2.2.3 Certifier, valider, imposer

Il est bien difficile de prétendre que le monde va devenir parfait et que l'ensemble des données sera toujours produit sous une forme « académique ». Il faut donc se donner les moyens de pouvoir les évaluer, d'une part car elles sont nombreuses actuellement, et d'autre part car elles ne devraient pas disparaître aussi facilement.

Ces données produites sans spécification initiale et sans contrôle qualité sont très nombreuses au sein des dynamiques régionales. On rappelle que ces données concernent une thématique ou un métier précis et intéressent forcément de nombreux acteurs locaux. Elles sont destinées à être partagées et devraient offrir des conditions minimales de qualité et de conformité. On sait qu'il n'en est rien. Certaines sont utilisées faute de mieux tant qu'elles répondent à certains besoins, d'autres non. Dans de telles conditions :

- ➔ Faut-il refuser de mettre sur la place publique de telles données, et les saisir de nouveau proprement ?
- ➔ Faut-il mettre en place de nouveaux outils, nouvelles méthodes ou nouveaux critères pour évaluer la qualité des lots de données et pour la communiquer ?
- ➔ Faut-il légiférer pour imposer un niveau de qualité minimum ?

Toutes ces questions offrent forcément de nombreuses réponses. En revanche, il apparaît clairement que dans la mesure où la qualité des données géographiques reste un sujet complexe à maîtriser par l'ensemble des acteurs, qu'elle est cependant incontournable pour fiabiliser les études et les prises de décisions qui en dépendent, que l'essor actuel de la géomatique voit poindre, sur la place publique, pléthore de données, il faut absolument disposer de moyens simples, capables de décrire cette qualité et les imposer sous peine de ne pouvoir diffuser ces données, et pire, de devoir les jeter.

Une règle, un règlement, une obligation ne servent que si l'on dispose d'un moyen de contrôle. Ceci introduit l'idée de validation ou de certification et d'organisme capable de les imposer. Ce principe n'a, bien sûr, pas pour objectif d'interdire la publication de données, ce qui irait totalement à l'encontre de l'engouement qu'on voit naître aujourd'hui autour des données collaboratives et de l'usage croissant de la géomatique. Son but serait avant tout de contraindre tout producteur de données qui désire les diffuser ou qui y est obligé à les accompagner systématiquement des informations nécessaires pour traduire leur qualité voire leurs usages possibles.

Aujourd'hui, au niveau des données de l'annexe I, la directive INSPIRE impose certaines métadonnées et quelques critères de qualité qui doivent accompagner la diffusion de tout lot de données. Cependant, un groupe de travail s'est mis en place

au sein de la commission européenne sur ce sujet. Les réflexions menées et les décisions prises pourront alors s'étendre au delà des données environnementales.

Il faut pour cela :

■ Définir les critères qualité qui doivent au minimum accompagner tout lot de données destiné à être diffusé.

■ Mettre en place les moyens légaux devant accompagner cette démarche.

Ce qui précède, montre à quel point l'usage de l'information géographique, que ce soit dans un contexte professionnel ou grand public, nécessite de précautions pour répondre à des besoins précis. Cela implique des normes strictes, le respect rigoureux de celles-ci, des moyens tout aussi rigoureux pour communiquer, des compétences reconnues pour produire les données, évaluer leur qualité et saisir les métadonnées. Dans un contexte aussi hétérogène que celui de la géomatique, où experts, utilisateurs occasionnels et grand public se côtoient, où chacun peut également devenir un producteur, il paraît impossible de proposer des méthodes universelles et utilisables par l'ensemble des acteurs.

On sait également à quel point la qualité ou plutôt la non-qualité des données peut avoir comme incidence grave, et impliquer jusqu'à la vie des individus qui se risquent à les utiliser sans contrôle préalable. Dans ce cas, la notion de responsabilité est mise en cause.

Ces constatations militent pour que le sujet de la qualité des données soit traité avec la plus grande attention, et pour la mise en place de méthodes de validation, voire de certification des données appliquées par un organisme d'expertise professionnelle reconnue. Cette expertise peut également être utile à différents niveaux comme l'aide à la rédaction de cahiers des charges ou de spécifications, la réalisation des contrôles qualité ou l'assistance à la saisie des métadonnées.

La géomatique est un outil d'experts mis dans les mains d'utilisateurs de tous niveaux. Les organismes ministériels et collectivités territoriales ont depuis longtemps mis en place des structures d'aide et d'accompagnement pour leurs services.

■ Il semble opportun d'étudier et d'envisager une structure nationale destinée à assister les utilisateurs et à certifier les données.

Dans ce qui précède, il est essentiellement question de validation et de certification de la donnée. Une autre démarche pourrait consister à valider ou certifier l'organisme responsable de la production de la donnée, voire le processus seulement. Ce qui aurait pour conséquence d'être assuré que la donnée est produite selon les normes, que les contrôles qualité sont effectués et les métadonnées et rapports qualités proposés.

3.3 Des compétences et une sensibilité des utilisateurs à développer

La complexité des normes et la difficulté de définir des spécifications préalables à toute production de données métier soulignent avant tout un manque de « professionnalisme » de certains utilisateurs qui ne maîtrisent pas l'ensemble de ces concepts. La géomatique est devenue un outil incontournable pour bon nombre de métiers qui manipulent au quotidien des données spatiales. En revanche, elle s'appuie sur des connaissances indispensables pour être pratiquée de façon optimale. Les concepts liés à la qualité des données nécessitent des connaissances accrues

et abordent tous les domaines techniques qui font parfois peur (géométrie, sémantique, statistique, techniques de saisie des données, informatique, etc.).

En dehors des formations initiales dispensées au sein des écoles et des universités spécialisées pour former des professionnels de la géomatique, peu d'écoles d'ingénieurs ou de techniciens abordent cette matière au cours de leur enseignement, alors qu'ils représentent une masse importante de futurs utilisateurs. Il est regrettable qu'à l'instar de l'informatique, la géomatique ne soit pas encore considérée comme un outil précieux d'analyse et de connaissance. Il devient indispensable de développer des formations complémentaires adaptées au monde des utilisateurs dont le métier premier n'est pas la géomatique mais l'utilisent comme un outil au sein de leur domaine de compétence.

Un autre aspect qui mérite d'être exploré est **d'accroître la sensibilisation des utilisateurs** autour des problèmes de qualité de leurs données. Il faut leur dire qu'elles peuvent être partiellement fausses ou entachées d'erreur. La notion d'utilisateur est ici prise dans sa globalité qui va du technicien qui manipule des données dans le cadre de son métier à l'utilisateur qui participe à la saisie des données sur un site collaboratif ou prépare une randonnée à l'aide des données disponibles sur internet.

Développer des actions de sensibilisation au sein des ministères, des collectivités territoriales, des organismes comme le CNIG ou des associations comme l'AFIGÉO

Pour un certain nombre d'entre eux, on peut imaginer que cette sensibilisation peut être le résultat naturel de la pénétration croissante des données spatiales dans la vie quotidienne des gens qui perçoivent que les données géographiques ne sont pas parfaites, simplement en regardant les lieux qu'ils connaissent bien sur des supports comme le GéoPortail, Google Earth ou Google Maps. Ils doivent naturellement intégrer la notion d'incertitude (voire accepter le risque) en utilisant les données, surtout lorsque la décision à prendre n'est pas critique [Devillers – 2010].

Il serait intéressant, d'attirer l'attention des utilisateurs sur les conséquences potentielles de l'utilisation de lots de données qui ne correspondent pas à leurs besoins, notamment en termes de qualité.

Ces réflexions conduisent naturellement à la notion d'incertitude qui entoure chaque donnée. Une forme quantifiée ou décrite de l'incertitude, au sein des métadonnées devrait pouvoir qualifier chaque donnée et permettre ainsi de mieux appréhender quantitativement l'impact de la qualité des données sur certains types de décisions à prendre, pour finalement mieux comprendre l'usage final des données.

Des études approfondies doivent être menées sur les moyens d'informer les utilisateurs des imperfections des données.

Un certain nombre d'études ont été réalisées sur les moyens de visualiser, et, plus généralement de communiquer les informations liées à l'incertitude. Cependant, aucune étude n'a clairement migré du secteur universitaire vers les utilisateurs et les milieux professionnels. Les chercheurs pourraient peut-être explorer d'autres manières d'aborder le problème et examiner de plus près les besoins des utilisateurs, au lieu de se concentrer sur les informations disponibles qui pourraient être communiquées [Devillers – 2010].

3.4 Un monde en perpétuelle évolution

On a pu constater, au cours des dix dernières années, une évolution du paysage dans lequel les données spatiales interviennent. Une forte démocratisation et une utilisation grand public de l'ensemble de ces données s'est produite, résultant entre autres de l'adoption et de l'évolution de l'internet.

De nombreux utilisateurs sans connaissance particulière en géomatique, utilisent des applications de cartographie web, comme Google Earth, Google Maps ou GéoPortail, pour toutes sortes de tâches imprévisibles, l'utilisation du GPS pour des déplacements motorisés ou pédestres, des sites d'aide à la gestion des itinéraires développent une nouvelle culture de l'utilisation accrue de l'IG.

Dans la nouvelle ère du Web 2.0, où chacun peut contribuer au contenu d'un site web :

- ➔ Comment peut-on évaluer la qualité des données produites par un utilisateur sur internet ?
- ➔ La cartographie doit-elle être limitée aux cartographes professionnels ou y a-t-il quelque chose de positif à ouvrir les portes au public qui désire y contribuer ?
- ➔ Comment peut-on évaluer la qualité d'un seul ensemble de données potentiellement produites et mises à jour par des milliers d'internautes ?

Certaines organisations de saisie collaborative comme OpenStreetMap prétendent avoir mis au point des procédures fiables pour répondre à ces questions en dehors des chaînes de production traditionnelle et proposent des données qui semblent répondre à de nombreux besoins.

Parallèlement, les organismes nationaux comme l'IGN s'interrogent pour offrir un accès plus facile à leurs données et imaginent de vrais échanges entre producteurs et utilisateurs pour produire des référentiels co-produits et largement partagés.

Ce paysage changeant pose également de plus en plus de questions juridiques, notamment au niveau de la responsabilité. Qui est responsable si quelqu'un utilise des données géographiques de mauvaise qualité et cause un préjudice ? Est-ce le producteur des données qui n'a pas fait un travail parfait ou qui n'a pas suffisamment documenté la qualité de ses données ? Est-ce le distributeur des données ? Est-ce l'utilisateur qui a pris des risques en utilisant les données disponibles et en ignorant leur qualité ? Ou est-ce une responsabilité partagée ? L'affaire devient encore plus complexe lorsqu'il s'agit d'agrégats de données ou de données produites au sein d'une démarche collaborative.

Le réseau canadien de recherche GEOIDE travaille actuellement sur ce vaste sujet et l'on peut espérer que la France et l'Europe feront de même. On a commencé à se poser des questions purement techniques dont les réponses ne sont pas particulièrement satisfaisantes avant d'aborder le problème juridique de la responsabilité. On peut s'interroger s'il ne serait pas préférable de partir du juridique pour déboucher sur des solutions techniques simples et en parfaite adéquation avec le niveau de risque et de responsabilités encourus.

3.5 Les métadonnées perçues comme un échec partiel

Hormis les rapports spécifiques prônés par la norme ISO 19114 et destinés à décrire la qualité d'un lot de données, les métadonnées restent, pratiquement, le

seul vecteur de communication. La publication de métadonnées demeure donc la principale méthode qui a été utilisée par les producteurs de données, au cours des deux dernières décennies, pour diffuser la qualité des données spatiales aux utilisateurs, qu'ils soient professionnels ou non. On peut déplorer que les normes n'aient pas de caractère obligatoire et que l'on puisse diffuser des données géographiques sans aucune référence à leur qualité.

En revanche, la directive INSPIRE va imposer un certain nombre de critères qualité pour la diffusion des données. Le tableau ci-dessous montre les critères retenus pour les données déjà spécifiées de l'annexe 1.

Légende du tableau :

AD : Adresses

UA : Unités administratives

PC : Parcelles cadastrales

NG : Noms géographiques

HY : Hydrographie

SP : Sites protégés

RT : Réseaux de transport

X : mesure de qualité utilisée dans les sous-critères (ISO 19138) / nombre de mesures qualité à réaliser

Critère qualité	Sous-critère	AD	UA	PC	NG	HY	SP	RT
Exhaustivité	Commission	X	X			X	X	X/2
	Omission	X	X	X	X	X	X	X
Cohérence logique	Cohérence conceptuelle	X	X			X		X
	Cohérence de domaine	X				X		X
	Cohérence de format							X
	Cohérence topologique		X/2	X/3		X/9		X/6
Précision géométrique	Précision absolue	X/2	X	X	X	X	X/2	X
	Précision relative					X		
Précision temporelle	Cohérence temporelle	X						
Précision thématique	Exactitude de classification							X
	Exactitude des attributs qualitatifs	X				X		X
	Précision des attributs quantitatifs					X		
Maintenance		X	X	X	X	X	X	X

Cette absence d'information sur la qualité des données du producteur vers l'utilisateur peut être perçue comme un échec partiel dans la mesure où elle ne répond absolument pas aux attentes.

Générer des métadonnées est une tâche qui prend du temps et qui s'est révélée avoir un impact assez limité sur la capacité des utilisateurs à comprendre les utilisations possibles des données. Les utilisations réussies de métadonnées demeurent assez basiques et sont principalement liées à l'utilisation de mots clés relativement simples (extension spatiale et temporelle, thème, etc.). Logiquement, les utilisateurs sont censés comprendre les caractéristiques d'un ensemble de données et l'étendue de son potentiel d'utilisation au travers de ses métadonnées. Malheureusement, il existe toujours un écart entre l'évaluation de la qualité que les experts de l'information géographique peuvent produire et ce que les utilisateurs peuvent comprendre et traduire en termes d'utilisations possibles. [Devillers – 2010 et Lex Comber - 2010]

La complexité de décrire la qualité est telle qu'il est difficile de documenter les métadonnées en se limitant à une instruction simple. On peut se demander s'il faut documenter au minimum à l'aide de métadonnées limitées mais totalement exploitables et compréhensibles pour permettre aux utilisateurs de lire et aux producteurs de renseigner ou s'il faut aller dans plus de détails et décrire plus finement et plus précisément. Cette seconde solution nécessite du temps, des compétences et n'est peut-être pas la solution la plus efficace dans le contexte actuel du développement de l'information géographique. La quasi absence de métadonnées de qualité faisant référence aux critères de la norme ISO 19113 en est l'illustration même.

Pour cette raison, l'équilibre entre ce qui devrait ou ne devrait pas apparaître dans les métadonnées n'est toujours pas résolu et peut varier en fonction des profils des utilisateurs et des usages.

- ➔ Les différents champs des métadonnées sont-ils pertinents dans le nouveau contexte d'une démocratisation de l'information géographique ?
- ➔ Y a-t-il des omissions importantes ?
- ➔ Y a-t-il des possibilités de richesse supplémentaire fournie par les récentes innovations en sciences de l'information ?
- ➔ Est-ce que les producteurs de données seront toujours en mesure de suivre la production des métadonnées selon les productions croissantes des données ?

Pour conclure sur ce sujet, on peut donc s'interroger sur les capacités des métadonnées à remplir leur rôle de documentation adéquate de la donnée et déplorer l'absence quasi récurrente de métadonnées de qualité, aujourd'hui, chez les producteurs de données géographiques.

Enfin, de nombreux utilisateurs réclament que la métadonnée soit plus directement liée à la donnée elle-même. Métadonnées et données sont, dans la plupart des cas gérées par des outils séparés ce qui ne facilite pas la consultation des métadonnées lors d'une session de travail et encore moins leur saisie dans une phase de production.

- ➔ Métadonnées et données doivent être plus imbriquées, voire intégrées dans un modèle commun et gérées par les mêmes outils.

Exemple de la « Généalogie »

La « Généalogie » est un critère de la norme ISO 19113 qui ne sera pas retenu dans la nouvelle norme ISO 19157. En revanche, elle est totalement présente en tant que métadonnée dans la norme ISO 19115. Cette métadonnée est obligatoire dans INSPIRE pour la quasi totalité des thèmes. On sait que la connaissance de l'historique de la production d'une donnée permet d'en déduire de nombreuses infor-

mations sur la qualité finale. En l'absence de contrôle qualité et de métadonnée de qualité, la généalogie doit pouvoir fournir une idée de la valeur qualitative des données. Elle exige, cependant, de la rigueur et des compétences indispensables pour la description des différents processus mis en œuvre.

Une investigation a été réalisée sur certaines fiches de métadonnées du Géocatalogue. Elle nous a montré combien cette métadonnée était très mal renseignée voire totalement « hors sujet ». C'est en règle générale un « fourre-tout » qui ne remplit que très rarement ses fonctions.

Il y a véritablement urgence à repenser cette métadonnée, à la structurer pour permettre de mieux guider les rédacteurs lors de la saisie. En règle générale, toutes les métadonnées formatées en texte libre sont difficilement exploitables et laissent trop libre-cours à l'interprétation.

3.6 Une terminologie à redéfinir

Une des raisons probables de l'échec partiel attribué aux métadonnées résulte, en partie, de l'absence d'une terminologie couramment acceptée dans le domaine de la qualité des données géographiques. L'utilisation de termes aussi variés au sein de la communauté géomatique peut être déroutante pour identifier les différents concepts, souvent complexes, exprimant la même idée mais désignés par des mots différents.

Ce sujet purement sémantique demeure donc particulièrement compliqué, et l'on peut regretter que les organismes de normalisation ne recherchent pas forcément la convergence des notions et de leurs définitions.

Prenons par exemple, la notion de « **Résolution spatiale** » que la directive INSPIRE impose et qui reste très difficile à interpréter, notamment pour les données vecteur.

Définition de la norme ISO 19115 : facteur qui fournit une idée, un ordre de grandeur de la densité spatiale des informations contenues dans un jeu de données.

Définition INSPIRE : 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 (normalement 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). Une échelle équivalente est généralement exprimée sous la forme d'une valeur entière correspondant au dénominateur de l'échelle. Une distance de résolution est exprimée sous forme de valeur numérique associée à une unité de longueur.

Ainsi, pour INSPIRE, cette grandeur est exprimée soit par une échelle pour les données de type vecteur, soit par une distance pour les données de type raster. Quant à la norme ISO, elle définit la résolution spatiale comme « une densité spatiale des informations contenues dans un jeu de données », ce qui est, d'une part, complètement différent de la définition proposée par INSPIRE et d'autre part, elle laisse libre cours à une interprétation complètement floue relayée par les mots « idée » et « ordre de grandeur » qui vont totalement à l'encontre d'un besoin de rigueur pour ce type de métadonnées. La notion d'échelle proposée par INSPIRE pour qualifier la résolution spatiale d'un lot de données vecteur est également très subjective et sujette à interprétation.

Dans cet exemple, on constate plusieurs dysfonctionnements : les définitions d'une même notion sont différentes, elles sont totalement discutables, trop propices à interprétation et peu rigoureuses.

Il serait opportun que les experts travaillant au sein des organismes de normalisation et des instances internationales assistés par des volontés nationales collaborent afin d'uniformiser la variété des termes et des définitions.

3.7 La qualité : à la frontière entre plusieurs cultures

3.7.1 Entre chercheurs et utilisateurs

La géomatique est une science nouvelle en perpétuelle évolution qui fait l'objet de nombreuses recherches et notamment autour de la qualité des données. Parallèlement, de multiples applications se développent, tant professionnelles que grand public, où les besoins de qualité s'avèrent de plus en plus prégnants. Des travaux sont réalisés et des études menées dans le monde de la recherche mais restent, très souvent, des exercices théoriques et académiques et n'atteignent pas le monde opérationnel.

Du côté des utilisateurs, les besoins ont parfois du mal à être clairement exprimés. Une assistance, des guides, des méthodes appropriées, des outils seraient certainement très utiles sur de nombreux sujets comme, par exemple, la saisie des métadonnées ou leur exploitation, la mesure de certains critères qualité ou de mesures spécifiques, la rédaction de cahiers des charges ou de spécifications ou tout simplement une méthode claire et simple pour aider à exprimer des besoins...

Il devient donc complètement opportun et urgent que ces deux mondes communiquent de plus en plus pour aboutir à des outils, des méthodes opérationnelles ou des guides qui répondent de façon pragmatique et ciblée aux besoins exprimés.

Il convient de souligner la démarche canadienne qui, grâce à son réseau de recherche GEOIDE, a mis en place, en 2009, un programme de recherche intitulé : « Protection du public et diffusion éthique des données géospatiales : aspects sociaux et légaux ». Ce programme de recherche vise à proposer des solutions concrètes aux problématiques liées à la qualité des données géographiques et à améliorer les principes de gestion visant à mieux encadrer la production, la diffusion et l'utilisation responsables des données. Outre ses objectifs très encourageants, il instaure un dialogue constructif entre le monde de la recherche et celui des utilisateurs qui participent à ce programme.

Les deux communautés françaises de chercheurs et d'utilisateurs pourraient s'inspirer de l'exemple canadien pour communiquer sur le sujet de la qualité.

3.7.2 Entre juristes et techniciens

Les aspects juridiques et techniques de la qualité des données géographiques demeurent totalement imbriqués et dépendants. Là également, on peut rappeler le programme de recherche du réseau GEOIDE qui l'a complètement intégrée. Ce présent rapport ne fait pas état des aspects juridiques liés à la qualité des données, mais il faudra en tenir compte très rapidement.

Si les juristes et les géomaticiens se rencontrent rarement, il faut tout de même rappeler l'initiative du Certu avec l'aide d'un cabinet d'avocats à produire un guide juridique, « Échanger des données localisées » et mettre en avant le colloque du

29 septembre 2009 de Montpellier (dimensions juridiques de l'information géospatiale en aménagement du territoire) organisé par l'Ordre des géomètres-experts, la faculté de droit de l'université Montpellier 1, Le département des sciences géomatiques de l'université Laval de Québec, le réseau canadien GEOIDE, sous l'initiative de Armelle Verdier-Maillot. Cette rencontre a d'abord le domaine de l'information géographique sous l'angle strictement juridique et a permis la rencontre juristes - praticiens et enseignants – spécialisés, chacun dans un domaine identifié du droit d'un côté et de la géomatique de l'autre. Une telle rencontre, qui dépassait les frontières, demeure une première. Il faut espérer qu'elle n'était qu'une introduction à une longue collaboration qui s'annonce.

La problématique juridique et celle de la responsabilité des producteurs et diffuseurs de données n'ont pas été abordées dans ce rapport, mais une suite sera proposée prochainement. C'est un point d'une grande importance et qu'il ne faut pas démarquer des aspects techniques. Quand Armelle Verdier-Maillot reviendra du Québec nous aborderons ensemble ce sujet.

Il est impératif que juristes et géomaticiens continuent de travailler ensemble sur le sujet de la qualité des données.

3.7.3 Entre professionnels et grand public

On peut, sans être trop caricatural, tenter de diviser la communauté géomatique en quatre classes :

- les chercheurs,
- les professionnels qui regroupent les producteurs de données, les éditeurs de logiciels et de matériels, les géomètres experts et les consultants,
- les utilisateurs qui pratiquent la géomatique comme un outil d'analyse dans le cadre de leur métier et produisent également des données thématiques,
- le grand public et quelques utilisateurs occasionnels qui utilisent essentiellement le GPS ou les outils mis à leur disposition sur internet.

Comme nous l'avons déjà souligné, les acquis et les niveaux de compétence sont très hétérogènes selon les acteurs. On a vu également que les outils existants autour de la qualité étaient produits par des professionnels et restaient relativement mal adaptés à un contexte non expert. Des données sont produites par des producteurs professionnels, d'autres par des utilisateurs pour leurs propres besoins et des données collaboratives par le grand public. Les qualités sont relativement bien maîtrisées pour certaines et moins bien pour d'autres mais les usages et les besoins sont très variés et chaque type de données semble trouver sa place.

- ➔ Devant cette diversité d'acteurs, d'usages et de qualité de données, on ne peut que s'interroger, sur les méthodes à mettre en place ?
- ➔ Faut-il appliquer les mêmes méthodes et les mêmes outils pour chaque type de données, d'usage et de besoin ?
- ➔ Faut-il faire intervenir d'autres critères comme par exemple la notion de risque, de responsabilité, la mise en place d'indicateurs ou d'autres méthodes qui restent à trouver pour évaluer la qualité, relevant dans certains cas de qualité interne et dans d'autres de qualité externe, voire des deux ?

4. Vers une nouvelle doctrine de la qualité

Devant la complexité de maîtriser les outils normatifs, que ce soit au niveau des mesures de qualité ou de la difficulté à renseigner et exploiter les métadonnées, il est certainement temps de repenser et de proposer des méthodes complémentaires et simples à mettre en œuvre. Il n'est pas question de remettre en cause ces outils mais il faut travailler sur leur déploiement et surtout sur une interface qui permettrait de faire le lien entre les producteurs dotés des outils complexes et les utilisateurs qui demandent plus de simplicité.

Sans prétendre à l'exhaustivité, nous pouvons retenir quelques pistes innovantes ayant fait l'objet, pour certaines, de publications et qu'il pourrait être intéressant d'approfondir, voire d'expérimenter.

4.1 Le tableau de bord et les indicateurs de qualité

Cette méthode¹ permet de traduire des métadonnées et des informations de qualité stockées en général dans des fichiers textes externes aux données, en des indicateurs plus facilement interprétables intégrés dans l'interface du SIG dans le but de réduire les risques de mauvaises utilisations. La qualité est ici définie comme « l'adéquation à l'utilisation ». L'information fournie aux utilisateurs compare leurs attentes aux spécifications internes des données. L'information sur la qualité est communiquée sous forme d'indicateurs de statut ou de risque que l'utilisateur peut sélectionner, modifier et consulter à différents niveaux de détails. Cette méthode nécessite le développement d'un outil basé sur trois technologies. Il n'existe aujourd'hui qu'un prototype (université Laval).

C'est très certainement autour de ce genre d'outil qu'il faut envisager de développer une interface entre producteurs et utilisateurs qui n'emploient pas le même langage et pour traduire des critères complexes en des indicateurs simples et compréhensibles par l'ensemble de la communauté des utilisateurs.

Parmi les indicateurs possibles, on peut imaginer différentes solutions très simples comme le smiley, des représentations comme les feux de circulations, des compteurs, des notes, etc. L'intérêt d'un tel système réside dans sa capacité à le paramétrer selon les compétences et les besoins de l'utilisateur.

4.2 Le « geotagging » ou Post-It® géographique

On désigne, sous ces anglicismes², la possibilité de marquer un emplacement géographique en y laissant un commentaire ou une évaluation. La société Arx iT (<http://www.arxit.com>) a développé ce type d'application dite de « geotagging » en utilisant un moteur de confiance pour évaluer la pertinence des commentaires ou des évaluations. Cette solution permet à chaque utilisateur de déterminer la fiabilité des tags qui sont déposés. Ce moteur de confiance électronique est capable de gérer et de renseigner l'utilisateur sur la pertinence d'une information dans un contexte particulier. Concrètement, les moteurs de confiance analysent les

1 Indicateurs de qualité pour réduire les risques de mauvaise utilisation des données géospatiales – Rodolphe Devillers, Yvan Bédard, Marc Gervais – Revue internationale de géomatique. Volume 14 – N° 1/2004

2 Le « geotagging » en toute confiance – Géomatique Expert – N° 71 – Octobre-Novembre 2009

commentaires que les utilisateurs déposent puis créent des communautés virtuelles de personnes partageant les mêmes goûts et intérêts. Ainsi, plus un utilisateur participe au système, plus il est reconnu par les autres participants et plus l'information qu'il reçoit en retour est pertinente.

Une telle démarche pourrait être employée autour de l'utilisabilité des données géographiques ou de l'adéquation d'une donnée particulière à une utilisation possible en fonction de l'expérience acquise par les différents utilisateurs.

4.3 Les réseaux sociaux professionnels

On sait la capacité des réseaux sociaux à rassembler et à communiquer, et qu'aujourd'hui, près de la moitié des internautes, dans le monde, accède à l'information via ces réseaux. Ce type de communauté occupe une place croissante du web. Face à un tel engouement, on ne peut pas ne pas s'interroger sur une utilisation raisonnée d'un tel phénomène. Twitter se présente aujourd'hui comme un « flux continu de conversations que l'on ne stocke plus et qui nous permettent d'accéder à l'information et surtout d'échanger avec les autres ». Au travers de ces réseaux, débute l'ère de la **recommandation** qui est une conséquence de cet aspect conversationnel où l'on « recommande » : la lecture d'un livre, le dernier film, l'achat d'un produit, etc. La recommandation prime très certainement sur la publicité traditionnelle car elle provient d'une personne du même réseau, donc plus proche, même si on ne la connaît pas, et revêt à ce titre plus d'importance qu'un message publicitaire forcément intéressé et moins objectif.

Associé à un géoportail qui permet de stocker certaines informations, comme l'est « GeoRezo » en France, les capacités offertes par les réseaux sociaux utilisés dans un contexte professionnel présentent un intérêt évident pour permettre aux utilisateurs d'échanger leurs expériences autour des usages et des utilisations des données et peuvent, en partie, résoudre certains aspects de la qualité externe.

4.4 Les spécifications initiales et l'expression des besoins

Les spécifications d'un lot de données constituent une condition importante et indispensable pour exprimer une évaluation de la qualité. C'est l'essence même de l'expression de la qualité interne.

Pour la qualité externe, la difficulté provient, d'abord, à exprimer clairement un besoin et à le comparer, ensuite, à des lots de données existants pour mesurer son niveau d'adéquation.

Spécifier une base de données et exprimer un besoin en termes de système d'information sont très proches en termes de finalité. La norme ISO 19131 constitue une aide à la spécification de contenu informationnel. On pourrait très bien imaginer une norme qui aide à exprimer et décrire un besoin sur les mêmes principes. Il suffirait alors de comparer, à l'aide d'outils spécifiques, les spécifications des lots de données pressentis aux spécifications du besoin exprimé pour évaluer un niveau d'adéquation.

5. Conclusion

Cette réflexion menée sur les dysfonctionnements constatés et sur la marge de progrès possible autour de la qualité des données géographiques s'appuie sur de nombreux échanges avec d'autres acteurs de la communauté géomatique. Elle n'est pas exhaustive et n'aborde pas les aspects juridiques cependant si importants. Elle n'a pour seule finalité que de sensibiliser les différentes instances et groupements d'acteurs concernés à instaurer, autour de ce sujet, un débat le plus officiel possible débouchant sur une démarche pragmatique et opérationnelle.

Il sera certainement difficile de trouver une réponse à toutes les interrogations disséminées dans ce rapport. Le sujet principal est à la fois technique et juridique et concerne donc des domaines très éloignés. Le contexte d'un monde en perpétuelle évolution, l'engouement croissant pour l'utilisation de l'information géographique, l'aspect institutionnel des décisions à prendre et la directive INSPIRE qui nous pousse à trouver des solutions rapidement, nous incitent à prendre ce sujet à bras le corps en dissociant le court terme de réflexions plus longues à mener.

Le développement de l'information géographique croît continuellement et il devient donc urgent d'améliorer la qualité des données, ne serait-ce que pour améliorer la pertinence des décisions qu'elle offre. Communiquer cette qualité avec des moyens simples et lisibles est un impératif pour permettre son utilisation maîtrisée, raisonnée et responsable.

6. Bibliographie

Rodolphe Devillers, Alfred Stein, Yvan Bédard, Nicholas Chrisman, Peter Fisher and Wenzhong Shi, 30 years of research on Spatial Data Quality – Achievements, failures and opportunities, Saint-John's, 2009, 13 pages

Rodolphe Devillers, Yvan Bédard, Marc Gervais, Indicateurs de qualité pour réduire les risques de mauvaise utilisation des données géospatiales, Revue internationale de géomatique. Volume 14 – n° 1/2004, 25 pages

Le “geotagging” en toute confiance, Géomatique Expert - n° 71 – Octobre-Novembre 2009, 3 pages

Rodolphe Devillers, Robert Jeansoulin, Qualité de l'information géographique, Paris, Hermès, 2005, 350 pages

7. Table des matières

1.Introduction.....	6
2.De la théorie à la pratique.....	7
2.1 La qualité interne.....	7
2.2 La qualité externe.....	7
3.Dysfonctionnements perçus autour de la qualité des données	9
3.1 La qualité des référentiels géographiques.....	9
3.1.1 Le constat.....	9
3.1.2 Les améliorations possibles.....	10
3.2 La complexité des outils et des règles.....	11
3.2.1 Le constat.....	11
3.2.1.1 La réalité normative.....	11
3.2.1.2 Manque de spécifications initiales.....	13
3.2.2 Les améliorations possibles.....	13
3.2.2.1 Proposer un outillage plus simple.....	13
3.2.2.2 Apprendre à spécifier et à contrôler les données.....	14
3.2.2.3 Certifier, valider, imposer.....	15
3.3 Des compétences et une sensibilité des utilisateurs à développer.....	16
3.4 Un monde en perpétuelle évolution.....	18
3.5 Les métadonnées perçues comme un échec partiel.....	18
3.6 Une terminologie à redéfinir.....	21
3.7 La qualité : à la frontière entre plusieurs cultures.....	22
3.7.1 Entre chercheurs et utilisateurs.....	22
3.7.2 Entre juristes et techniciens.....	22
3.7.3 Entre professionnels et grand public.....	23
4.Vers une nouvelle doctrine de la qualité.....	24
4.1 Le tableau de bord et les indicateurs de qualité.....	24
4.2 Le « geotagging » ou Post-It® géographique.....	24
4.3 Les réseaux sociaux professionnels.....	25
4.4 Les spécifications initiales et l'expression des besoins.....	25
5.Conclusion.....	26
6.Bibliographie.....	27
7.Table des matières.....	28

© ministère de l'Écologie, du Développement durable, des Transports et du Logement

Centre d'Études sur les réseaux, les transports, l'urbanisme et les constructions publiques.

Toute reproduction intégrale ou partielle, faite sans le consentement du Certu est illégale (loi du 11 mars 1957). Cette reproduction par quelque procédé que ce soit, constituerait une contrefaçon sanctionnée par les articles 425 et suivants du code pénal.

Dépôt légal : 4^e trimestre 2010

ISSN : 1263-2570

ISRN : Certu/RE--10-14--FR

Certu

9, rue Juliette-Récamier

69456 Lyon cedex 06

(+33) (0) 4 72 74 59 59

Internet <http://www.certu.fr>

*Service technique placé sous l'autorité
du ministère chargé du Développement durable, des Trans-
ports et du logement,
le Certu (centre d'Études sur les réseaux, les transports,
l'urbanisme et les constructions publiques)
a pour mission de contribuer au développement
des connaissances et des savoir-faire et à leur diffusion
dans tous les domaines liés aux questions urbaines.
Partenaire des collectivités locales
et des professionnels publics et privés,
il est le lieu de référence où se développent
les professionnalismes au service de la cité.*

ISSN 1263-2570
ISRN Certu/RE--10-14--FR