Vasseure, Berengere, Robert Jeansoulin, Rodolphe Devillers, and A. U. Frank. "Evaluation De La Qualité Externe De L'information Géographique: Une Approche Ontologique." In *Qualité De L'information Géographique: Traité Igat*, edited by Rodolphe Devillers and Robert Jeansoulin, 285-301: Hermes Science, 2005.

Chapitre 15

Evaluation de la qualité externe de l'information géographique : une approche ontologique

15.1. Introduction

Les données géographiques proviennent de sources variées et il semble facile de les combiner par leur référence spatiale et de les utiliser pour analyser un problème précis et prendre une décision. Cependant, nous pouvons nous demander quelle est la valeur d'une décision basée sur des données dont la qualité est mal connue ou mal comprise par le décideur. Cela met en lumière le problème d'adéquation d'utilisation appelé aussi « *fitness for use* » ou « *qualité externe* » d'une application. Par conséquent, il est important d'aider le décideur à évaluer cette qualité externe de l'information. Or, il n'existe pas d'outil formel pour accompagner l'utilisateur dans sa démarche de sélection des données en accord avec ses attentes de qualité. D'après [HUN 01], « les utilisateurs voudraient avoir la possibilité technique de prendre l'information de qualité et de l'utiliser pour déterminer quel résultat proviendra de l'utilisation de ces données, les modèles, les opérations spatiales, avant que la tâche ne soit réellement entreprise. Actuellement, cela a été seulement exécuté de manière très limitée par quelques experts qualifiés, et cette fonctionnalité n'existe pas généralement

---

Chapitre rédigé par Bérengère VASSEUR, Robert JEANSOULIN, Rodolphe DEVILLERS et Andrew FRANK.

dans les progiciels commerciaux » (traduction libre). [FRA 98] suggère que la qualité des données doit être indépendante de la méthode de production, être opérationnelle pour les utilisateurs et utilisée dans une procédure formelle fournissant des résultats quantitatifs. Il existe donc un besoin pour mieux évaluer la qualité externe.

L'objectif de ce chapitre est de présenter une approche permettant d'évaluer la qualité externe de l'information géographique. Nous définissons tout d'abord les notions de qualité et d'ontologie. Nous présentons ensuite une méthode dite « ontologique » permettant de modéliser, évaluer et améliorer la qualité externe de l'information géographique. Nous illustrons notre démarche sur un exemple pratique pour comparer les besoins et données. Nous concluons ce chapitre en identifiant des perspectives de recherche.

15.2. Qualité et Ontologie

Nous présentons ici les notions de qualité et d'ontologie nous permettant de mieux appréhender l'évaluation de la qualité externe de l'information géographique.

15.2.1. *Notion de qualité et de qualité externe*

Deux définitions de la qualité des données peuvent être identifiées dans la littérature (voir chapitre 3 pour plus de détails). La première limite la qualité des données aux caractéristiques internes des données, découlant des méthodes de production des données (acquisition des données, modèles de données, etc.). Elle représente la différence entre les données produites et les données définies initialement (c'est-à-dire terrain nominal). Cette définition est souvent appelée « qualité interne » [AAL 98, AAL 99, DAS 03]. La seconde définition suit le concept de « *fitness for use* » [CHR 83, JUR 74, VER 99], correspondant au niveau d'adéquation existant entre les caractéristiques des données et les besoins des utilisateurs pour différents aspects (par exemple couverture spatiale et temporelle, actualité). Cette définition, souvent appelée « qualité externe » concerne donc l'aptitude des données à répondre aux besoins implicites ou explicites de l'utilisateur. Ce chapitre porte plus spécifiquement sur l'évaluation de la qualité externe.

Issue du domaine de la production industrielle, la notion de qualité externe a été introduite dans le domaine de l'information géographique numérique en 1982 par le comité fédéral américain des normes cartographiques digitales, reprise en 1985 dans un rapport consacré à la qualité cartographique numérique des données [MOE 87]. A l'heure actuelle, la définition de la qualité retenue par l'ISO correspond à la notion de qualité externe. Elle est définie comme étant « l'aptitude d'un ensemble de caractéristiques intrinsèques à satisfaire des exigences » [ISO 00].

Bien que le concept d'adéquation d'utilisation ait été adopté il y a plus de 20 ans comme définition de la qualité pour l'information géographique, il n'y a presque pas eu d'avancées depuis, portant sur le développement de méthodes pour évaluer l'adéquation de données pour une utilisation définie [VER 99]. Aussi, l'évaluation de cette adéquation reste dans les mains de l'utilisateur final, se faisant habituellement de manière intuitive, basée sur l'expérience de l'utilisateur, sur les conseils qu'il a reçu d'un expert ou sur les informations disponibles sur les caractéristiques des jeux de données. Aussi, les métadonnées (voir chapitre 12) sont le moyen habituellement utilisé pour aider l'utilisateur à évaluer cette adéquation d'utilisation. Les outils de type géorépertoire permettent aux usagers, par le biais de requêtes sur les métadonnées, de sélectionner des données adéquates à l'intérieur d'un ensemble de jeux de données disponibles (sélection des données pour des couvertures spatiale et temporelle d'intérêt, de la date de production des données, de leur format, etc.). En recherche, les travaux portant sur la visualisation de l'incertitude (voir chapitre 14) permettent d'aider les utilisateurs à évaluer l'adéquation d'utilisation. Cette dernière technique n'est toutefois pas encore implantée dans les SIG commerciaux. Plus récemment, plusieurs travaux ont visé à établir des approches plus formelles pour la définition de l'adéquation d'utilisation. [AGU 98] propose de diviser ces approches en deux catégories. La première, basée sur des normes (« *standard-based* »), compare l'incertitude intrinsèque aux données à un ensemble de normes reflétant des niveaux acceptables d'incertitude. La seconde, basée sur une étude de risque (« *risk-based* »), s'intéresse à l'évaluation de l'impact potentiel que peuvent avoir ces données incertaines sur les décisions allant être prises.

15.2.1.1. *« Standards-based »*

[FRA 98] présente un métamodèle permettant de fusionner les points de vues des producteurs et utilisateurs de données. Il critique l'utilité des métadonnées actuellement diffusées et prône des descriptions de la qualité plus indépendantes des méthodes de production, opérationnelles et quantitatives.

[VAS 03] présente une approche utilisant des ontologies pour formaliser les caractéristiques des données ainsi que les besoins des utilisateurs. Les ontologies fournissent alors deux modèles comparables (un formalisant les caractéristiques des données et l'autre les besoins des utilisateurs), facilitant alors les mesures de similarité entre ces deux ontologies (c'est-à-dire la qualité augmente avec la similarité). Cette approche est présentée en détail dans ce chapitre.

15.2.1.2. *« Risk-based »*

[AGU 98] explore le problème sous une autre optique. Au lieu de s'interroger sur la similarité *a priori* entre données et besoins, il analyse les risques potentiels *a posteriori* suite à l'utilisation de données pour une certaine application. Basés sur

des travaux en gestion du risque, il définit différentes étapes dans le processus : identification du risque (qu'est-ce qui peut aller mal et pourquoi), analyse du risque (probabilités et conséquences), évaluation du risque (quel est le risque acceptable), exposition au risque (quel est le niveau de risque réel), estimation du risque (comparaison entre le niveau de risque acceptable et l'exposition au risque) et réponse au risque (comment le risque peut être contrôlé). Cette approche implique une quantification des risques pouvant être difficile à estimer.

[DEB 01] propose une approche, basée sur le concept de « valeur de contrôle » permettant de comparer divers jeux de données alternatives pour une application. Cette approche, utilisant la technique des arbres de décision, nécessite une quantification de l'exactitude des données et une estimation des coûts potentiels qu'engendrerait une décision incorrecte basée sur ces données. Elle permet, grâce à une estimation quantitative des risques, de sélectionner le jeu de données qui va limiter les conséquences négatives faisant suite à la prise de décision.

15.2.2. *Complexité de l'évaluation de la qualité externe*

L'évaluation de la qualité externe semble provenir d'une simple comparaison entre les caractéristiques des données et les attentes de l'utilisateur et est souvent faite de manière intuitive [GRU 04]. Pourtant, cette évaluation est plus complexe qu'elle n'y paraît. En effet, la notion de qualité externe, liée à l'aptitude à satisfaire les besoins de l'utilisateur est évolutive, floue et complexe à mettre en œuvre. En premier lieu, nous considérons que la satisfaction des besoins de l'utilisateur est liée à un processus évolutif et convergeant vers une situation satisfaisante, où l'utilisateur modifie ses besoins en accord avec les données présentes, ou inversement cherche de nouvelles données. Cela rejoint la définition de la qualité du domaine de la production de biens, où la qualité vise à satisfaire le client dans une démarche d'amélioration continuelle [JUR 74]. De plus, la meilleure qualité est obtenue lorsque la différence entre l'état des données et les besoins est minimale. Si un utilisateur a besoin par exemple de données récentes, qu'entend-t-il par « récente » ? Où mettre la limite entre données « récentes » et « non récentes » ? L'utilisateur sera-t-il satisfait de données datant d'une semaine, un mois, un an ? Cette limite, en plus d'être floue, est contextuelle à chaque utilisateur mais également à chaque utilisation. Enfin, les caractéristiques des données et des besoins doivent pouvoir être comparées afin de pouvoir calculer une « utilité » des données. L'utilité constitue une mesure quantitative de la qualité externe, provenant de la comparaison entre les attentes des utilisateurs et ce que peuvent offrir les données disponibles [FRA 04]. Cela implique de disposer d'un référentiel commun de comparaison et il est souvent nécessaire d'effectuer des traductions et normalisations pour pouvoir être en mesure de comparer des caractéristiques diverses.

Nous définissons la qualité externe comme le degré de similarité existant entre les besoins des utilisateurs et les données, exprimés dans le même ensemble de formules logiques, que nous qualifions de référentiel, tout en s'intégrant dans un processus d'amélioration visant à satisfaire l'utilisateur.

15.2.3. *Notion d'ontologie*

La notion d'ontologie peut aider à confronter et à mesurer la similarité entre les besoins et les données, mais cela nécessite de formaliser les besoins des utilisateurs et les caractéristiques des données. Afin d'avoir des éléments comparables, les besoins et les caractéristiques sont formalisés dans des modèles rendus comparables par leur traduction dans un ensemble commun de formules logiques, censés représenter des ontologies éventuellement non comparables directement.

La réalité du monde et la connaissance qu'on en a sont des questions récurrentes de la philosophie. Le mot ontologie vient des racines grecques « *ontos* » et « *logos* », formant ainsi la science de « ce qui est ». Les philosophes ont étudié pendant des siècles l'existence, la connaissance et la description de l'être à travers ce concept d'ontologie. Il a été repris en informatique pour être défini comme la « spécification explicite d'une conceptualisation » [GRU 93], ou comme « l'ensemble des définitions des classes, relations, et autres objets d'intérêt (physiques ou logiques) dans un domaine d'intérêt » [NOY 01]. L'ontologie permet ainsi de partager un vocabulaire présentant la structure de l'information, de fournir une base de connaissance, un cadre commun de raisonnement échangeable et compréhensible par d'autres personnes non spécialistes du domaine. Il n'y a pas une manière unique de modéliser un domaine mais plusieurs alternatives qui dépendent de l'application.

Dans le domaine de l'information géographique, Mark définit une ontologie comme « la totalité des concepts, catégories, relations et processus géospatiaux » [MAR 02]. La publication d'une ontologie géospatiale fournit alors un modèle de raisonnement et permet d'intégrer différentes contraintes et de définir la sémantique et la terminologie d'une application géospatiale [FRA 03]. Hunter décrit une ontologie comme « une manière formelle de réaliser une description claire et concise des termes et des concepts que nous employons, afin que les autres puissent interpréter ce que nous faisons » [HUN 02]. La littérature identifie plusieurs niveaux d'abstraction dans les ontologies, s'apparentant aux niveaux existants dans le domaine de la modélisation des données, comme la représentation conceptuelle d'une réalité (objets, classes, etc.), la représentation logique, et la représentation physique. Aussi, pour la qualité externe, les différents niveaux de granularité peuvent être la description d'un problème et des données, la représentation conceptuelle des différents éléments importants de l'application et des données, les spécifications des besoins et des données.

Ainsi, en sciences de l'information, la notion d'ontologie semble très proche de celle de modèle (orienté-objet) associé à un vocabulaire partagé. Dans notre approche, nous nous intéressons à l'information géographique et aussi à la qualité de cette information, la référence au monde réel intervient donc à double titre : en tant que réalité partagée par l'utilisateur et le producteur de données, et en tant qu'horizon supposé d'une qualité parfaite. Cette référence au réel renvoie donc à l'ontologie, au sens philosophique, aussi bien qu'à l'ontologie d'application, au sens des sciences de l'information. Nous avons besoin de différents niveaux d'abstraction : un niveau de définition des entités utiles au problème considéré (une ontologie), un niveau de représentation de ces entités (modèle objet), et une traduction dans un formalisme adéquat (modèle relationnel).

15.3. Evaluation ontologique de la qualité externe

15.3.1. *Processus d'amélioration de la qualité*

La notion de qualité est associée à la notion de satisfaction du client et s'inscrit dans un processus d'amélioration continuelle [ISO 90, JUR 74]. En premier lieu, la question initiale du problème de l'application doit être spécifiée. Par la suite, les éléments disponibles, compatibles avec les attentes de l'utilisateur sont sélectionnés. Les besoins des utilisateurs sont progressivement atteints en revenant successivement à la conceptualisation initiale du problème et en reformulant le problème où en modifiant les données. Ce processus itératif converge vers une situation satisfaisante pour l'utilisateur, alliant l'ontologie du problème et des données et pouvant être représenté sous forme d'une matrice de qualité. La figure 15.1 illustre ce processus.

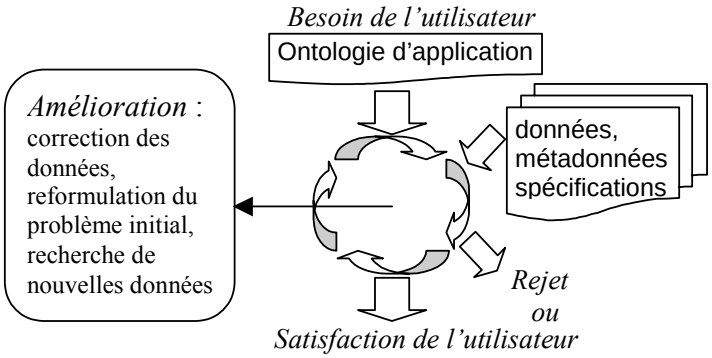

Figure 15.1. *Amélioration de la qualité - adapté de la Roue de Deming [ISO 00]*

15.3.2. *Intégration géosémantique*

Pour évaluer la qualité externe des données, on a besoin de capturer la connaissance de l'utilisateur (ontologie du problème) et des données (ontologie du produit) (voir figure 15.2). L'approche consistant à les plonger dans une ontologie d'application commune (référentiel) est appelée par Brodeur « espace géosémantique » [BRO 03]. Il est nécessaire d'effectuer une traduction des divers éléments dans le même référentiel. L'intégration géosémantique peut aussi être réalisée à un niveau plus général [COM 03, GUA 98, KOK 01, SOW 98]. Lorsque la traduction est effectuée, on peut comparer deux matrices de qualité « attendue » et « réelle » et fournir la matrice « de qualité de l'application », résultant de la comparaison (détails en section suivante). Cette représentation fournit une structure pour documenter, mesurer et communiquer la qualité de l'application [VAS 03].

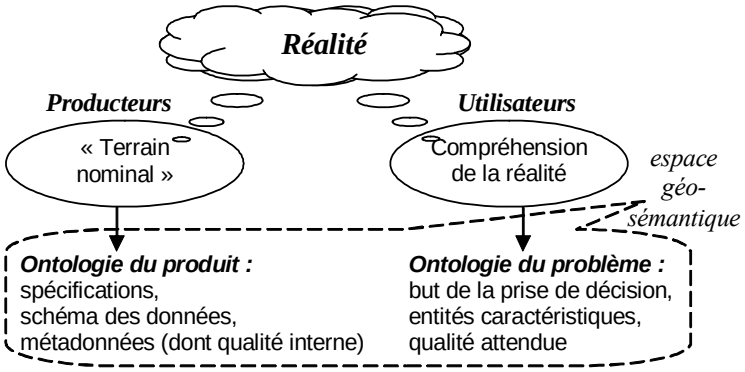

Figure 15.2. *Approche ontologique d'évaluation de la qualité externe*

15.3.3. *Etapes permettant d'évaluer la qualité externe*

Le processus d'amélioration de la qualité externe peut être modélisé en plusieurs étapes, tel qu'illustré sur la figure 15.3.

Les étapes de cette démarche sont :

– Conceptualiser la question posée et les hypothèses de travail dans une ontologie du problème. On n'explicite que ce qui est nécessaire au départ, incluant les besoins en qualité, on formalise et on construit la matrice de qualité attendue ;

– Examiner les catalogues de sources de données, les métadonnées des sources candidates et rechercher, au travers de leurs spécifications, en quoi leurs ontologies de produit (qui ont présidé à leur création et enregistrement) se rapprochent de l'ontologie du problème ; on construit la matrice de qualité interne ;

– Etablir la matrice de qualité de l'application qui évalue si les données sont compatibles avec la qualité attendue par l'utilisateur. Si c'est le cas, on effectue la traduction directe (4), sinon, on cherche une autre solution dans l'étape (5) ;

– Si les ontologies sont « ressemblantes », par exemple pour certains objets et attributs, il faut traduire la partie correspondante en une requête sur les données ; la sélection puis l'analyse de ces données permet d'améliorer éventuellement la matrice de qualité ;

– Sinon, il faut reformuler partiellement l'ontologie du problème en puisant dans les connaissances non explicites au départ, celles qui peuvent permettre de répondre indirectement aux entités manquantes, par une « chaîne d'inférences » entre les données et les questions, qui nous ramène à l'étape (1) ;

– La décision finale sur l'acceptation ou le rejet du résultat, est ainsi « éclairée » à chaque étape par une évaluation des limites tolérables de qualité, dans le contexte de l'adéquation à sa question initiale, avec une représentation de la qualité sous forme d'une matrice.

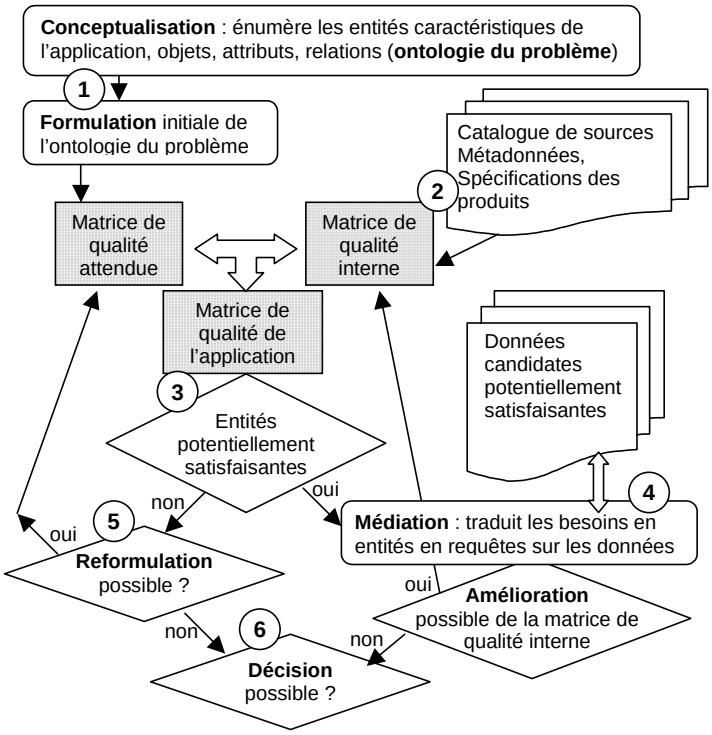

Figure 15.3. *Méthode d'amélioration de la qualité externe*

Cette procédure est bâtie sur deux roues de Deming :

– La boucle (1-2-3-5-1) est effectuée en premier, tant que les catalogues et métadonnées peuvent nous renseigner utilement sur les données disponibles candidates à une satisfaction espérée. Ce premier tri est rapide et gratuit car les informations sont (devraient être ?) disponibles sur Internet ;

– La boucle (2-3-4-2) possiblement imbriquée à nouveau avec la précédente, fait appel aux données elles-mêmes, donc achetées, et au coût d'un flot de données bien plus important (calcul d'agrégats sur de gros volumes), mais après le précédent tri qu'on peut espérer efficace : les données hors sujet, ou de trop mauvaise qualité, ou au contraire bien trop détaillées, auront été « filtrées » avant achat éventuel.

Les trois premières étapes de cette démarche sont approfondies : la spécification des besoins, les matrices de qualité et leur comparaison.

15.3.3.1. *Conceptualisation et formalisation initiale (étape 1)*

L'identification et la description des entités dont l'utilisateur a besoin pour répondre au problème qui lui est posé, sont un problème difficile. C'est une question de nature ontologique (« l'espace géosémantique »). De plus cette description doit s'accompagner d'une description de la qualité attendue pour chaque entité, en fonction de la décision à prendre. L'utilisateur doit posséder une expertise minimale dans son domaine (par exemple forestier, urbaniste ou même touriste) pour énumérer les entités utiles pour exprimer sa compréhension du problème et spécifier la qualité attendue pour chaque entité (par exemple précision spatiale métrique, bâtiments représentés par des polygones, etc.). Il faudrait à ce niveau un minimum d'outils d'aide à la formulation initiale de l'ontologie et à sa reformulation lors de la boucle, mais de tels outils sont encore du domaine de la recherche (le lecteur pourra se référer aux travaux de projets européens en cours, comme *OntoWeb, Web of Knowledge, Knowledge Web* ou autres projets équivalents).

15.3.3.2. *Matrices de qualité attendue et interne (étape 2)*

On construit deux matrices (voir figure 15.4) basées sur le même produit cartésien (voir ci-après) : une matrice représente la qualité interne des données (c'est-à-dire basée sur la qualité telle que mesurée ou documentée), une autre représente la qualité attendue (c'est-à-dire jugée satisfaisante par l'utilisateur dans le cadre de son application).

Ensuite on compare ces deux matrices, cellule par cellule, avec des mesures de similarité appropriées, et le résultat constitue une matrice dite « matrice de la qualité de l'application », qui identifie les entités pour lesquelles la qualité externe (c'est-à-dire écart entre données et besoins) est correcte, ou non.

Le produit cartésien support de ces trois matrices est constitué par :

– les entités de l'application présentes dans « l'ontologie du problème » construite pour la prise de décision *(Ci)* ;

– une liste d'éléments de qualité, parmi ceux identifiés par les standards ou des éléments nécessaires à l'utilisation (*Qi*).

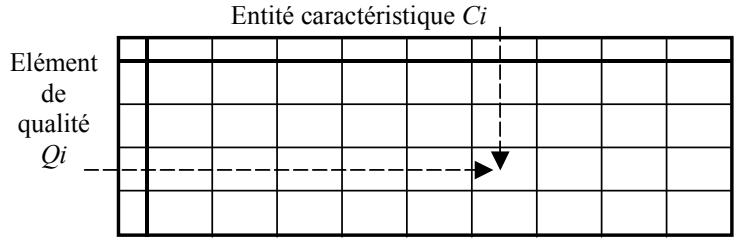

Figure 15.4. *Matrice de qualité*

15.3.3.2.1. Entités dans une ontologie d'application dite « de problème »

L'espace géosémantique cité précédemment peut se modéliser à l'aide d'entités caractéristiques *(Ci)* qui peuvent être des objets dans un état donné (une parcelle), des phénomènes dynamiques (un flux), des ensembles liés (un réseau), etc. Il n'existe pas de nomenclature unique pour toutes les entités d'une application. Comme le dit Hunter, « il n'y a pas de bonne ou de mauvaise description. Il y a juste des sens différents pour des buts différents, pour des gens différents » (traduction libre) [HUN 02]. On peut utiliser des vocabulaires issus de différentes techniques de modélisation (exemple entité-relation *versus* orienté-objet), on peut se baser sur les définitions de l'UML (*Unified Modeling Language*), utilisées par différents organismes internationaux dont l'ISO [UML 00]. Cette nomenclature a l'avantage de prendre en compte la dimension temporelle, avec les processus et les événements inhérents à certaines applications géospatiales. On peut en particulier s'intéresser à la description d'un état à un moment donné et à la description d'un processus continu (par exemple l'érosion), comme séquence de plusieurs états et de plusieurs matrices de qualité correspondant aux différents états [VAS 04].

15.3.3.2.2. Eléments de qualité

La norme ISO 19113 [ISO 02] définit 5 éléments et 15 sous-éléments quantitatifs pouvant aider à évaluer la qualité externe. Ces éléments concernent les paramètres d'*exhaustivité*, de *cohérence logique*, et d'*exactitude* (l'exactitude pouvant être de position, temporelle ou thématique). Par ailleurs, d'autres critères que ceux de la norme ISO 19113, sont définis dans la littérature comme rentrant dans l'évaluation de

la qualité externe. Par exemple, l'*actualité* traduit la différence entre la date de création du jeu de données et la date à laquelle il va être utilisé, la *couverture* permet de connaître le territoire (couverture spatiale) et la période (couverture temporelle) que les données décrivent, la *légitimité* permet d'évaluer la reconnaissance officielle et la portée légale d'une donnée (par exemple donnée produite par un organisme reconnu) et enfin l'*accessibilité* permet d'évaluer la facilité d'accès aux données [BED 95].

15.3.4. *Comparaison et utilité : exemple d'application en transport (étape 3)*

Afin d'illustrer l'évaluation de la comparaison issue de la matrice de qualité et du calcul d'utilité entre les besoins d'un problème et les données, nous présentons une application étudiée par l'institut technologique universitaire de Vienne dans le contexte du projet européen *REVIGIS* [FRA 04, GRU 04]. Il s'agit d'un problème d'aide à la navigation en transport urbain pour deux catégories d'utilisateurs : des touristes et des pompiers. Ils doivent se déplacer d'un point A à un point B et prendre la décision de tourner à droite ou à gauche en fonction des données disponibles.

15.3.4.1. *Ontologie du problème*

Les besoins et la prise de risque sont différents pour les deux groupes. Pour les pompiers, le but est d'atteindre une destination en ville le plus rapidement possible pour répondre à une urgence. La notion de « distance temps » est très importante. Elle évalue la durée nécessaire à un parcours. Les informations sur l'accessibilité au réseau des rues, les adresses et les emplacements des bornes d'incendies doivent être de très bonne qualité. Pour les touristes, le but est de se déplacer en ville pour découvrir l'espace d'un point de vue touristique, à pied ou en transport en commun, et la notion de « distance perçue » est privilégiée. Elle se mesure moins en termes de distance physique que par des possibilités de contact d'information ou de familiarité avec le lieu. Pour ces derniers, la qualité des informations concernant les points d'intérêts, réseau de transport en commun est très importante.

Une matrice de la qualité est faite pour chaque groupe d'utilisateur, associant 16 couples caractéristiques regroupés en thèmes (par exemple point d'intérêt, nom des rues, numéro des bâtiments) et 3 éléments de qualité (par exemple actualité, exhaustivité et exactitude). Un thème décrit une combinaison de plusieurs classes d'objets créant une unité complexe. Par exemple, les thèmes de la Base nationale de données topographiques (produite par Géomatique Canada) est effectivement un groupe de classes d'objets regroupées à cause de caractéristiques communes (par exemple le thème hydrographie pourra regrouper les rivières, lacs, etc.). Ces matrices indiquent la qualité désirée pour chaque type d'utilisateur et pour chaque couple contenu dans une cellule. La qualité indiquée pour une cellule indique la

zone de tolérance (elle est « nul » lorsque les cellules sont vides). Elle décrit une zone de valeurs acceptables correspondant à l'écart toléré entre la valeur idéale (dans la matrice de qualité réelle) et la valeur réelle (dans la matrice de qualité réelle) pour un couple donné. Par exemple, comme l'illustre le tableau 15.1, pour l'attribut « *points d'intérêt* », le groupe de touristes souhaitera obtenir les qualités avec les tolérances suivantes : actualité (≤ 2002), exhaustivité (≥ 95 %), et exactitude (≤ 1 mètre). En revanche, le groupe des pompiers, ne se base pas sur ces données pour prendre la décision de tourner et n'a donc aucune attente de qualité concernant ce thème, au contraire ce thème nuit à la visibilité des autres (≤ –50 %).

Points d'intérêts	**Qualité Attendue**	**Points d'intérêts**	**Qualité Attendue**
Touristes		**Pompiers**	
Actualité	2002	Actualité	*nul*
Exhaustivité	95 %	Exhaustivité	–50 %
Exactitude	1 mètre	Exactitude	*nul*

Tableau 15.1. *Qualité attendue pour la caractéristique « points d'intérêts »*

15.3.4.2. *Ontologie du produit*

Deux bases numériques sont présentes pour permettre aux deux groupes de prendre la décision de se diriger : la base de données numérique « multi-usage » (MPM, *Multi Purpose Map*) et la « base de données numérique de Vienne » (CMV, *City Map of Vienna*). Ces matrices associent les mêmes couples que dans les matrices de qualité désirée. Par exemple, l'attribut « points d'intérêt », comme présenté dans le tableau 15.2, est présent dans la base de donnée numérique CMV avec les qualités suivantes : actualité (1999), exhaustivité (90 %), et exactitude (0,5 mètre) mais cette caractéristique n'est pas présente dans la base de données MPM et les cases de la matrice correspondantes sont nulles.

Points d'intérêts	**Qualité Actuelle (MPM)**	**Qualité Actuelle (CMV)**
Actualité	*nul*	1999
Exhaustivité	*nul*	90 %
Exactitude	*nul*	0,5 m

Tableau 15.2. *Qualité actuelle pour la caractéristique « points d'intérêts »*

15.3.4.3. *Comparaison des données aux besoins et calcul d'utilité*

Les matrices de qualité réelles et attendues sont établies dans le même référentiel géosémantique. Aussi, il est donc possible d'effectuer le calcul d'utilité. Il s'effectue à travers différentes étapes de comparaison, normalisation et agrégation.

La comparaison mesure l'écart entre la valeur attendue et la valeur réelle pour chaque couple de cellules similaires <*a*, *a'*> comme l'illustre la figure 15.5. Les couples doivent être dans la même unité de mesure (par exemple jour *versus* mois, mètre *versus* cm) et il convient de modifier l'unité si cela n'est pas le cas. Il est également nécessaire de normaliser les unités pour chaque cellule de la matrice de qualité de l'application résultante pour obtenir une même unité <d(a_1), d(a_2), …d(a_n)>. Aussi, dans l'application de Vienne, le processus de normalisation a permis de traduire les valeurs de la matrice de qualité de l'application dans un référentiel commun, compris entre 0 et 100 % (ou 0 et 1), au lieu d'avoir des résultats exprimés en unité de pourcentage, mètre et année.

Le processus d'agrégation permet d'avoir une vision synthétique de la qualité. Cependant, il est contextuel aux données et au groupe d'utilisateurs et très sensible à la méthodologie utilisée. Aussi, l'utilisation des indicateurs doit être prise avec grande précaution puisqu'ils diffèrent selon les applications, les utilisateurs, la méthodologie. Les règles relatives aux opérations d'agrégation, le poids des différentes valeurs de la matrice, les différentes exceptions doivent êtres explicités. Par exemple, l'ISO 19114 [ISO 03a] décrit plusieurs possibilités pour agréger les données décrivant la qualité (ADQR, *Aggregated Data Quality Results*). Dans l'application de Vienne, l'agrégation provient d'une moyenne pondérée des valeurs des cellules de la matrice de qualité de l'application. Les résultats obtenus sont très significatif avec les résultats d'utilité suivants, illustrés dans le tableau 15.3 (nous ne détaillerons pas ici l'ensemble de ces règles explicitées dans [FRA 04] et [GRU 04]).

Utilité	**MPM**	**CMV**
Touristes	82 %	39 %
Pompiers	44 %	75 %

Tableau 15.3. *Utilité des bases de données pour chaque groupe d'utilisateurs*

A l'étude du tableau, nous constatons que la base de données numérique « MPM » est plus utile (c'est-à-dire qualité externe plus élevée) pour le groupe de touristes (82 %) tandis que la base de données « CMV » est plus intéressante pour le groupe de pompiers (75 %).

15.4. Discussion et conclusion

La méthodologie présentée permet de tendre vers un accord entre les besoins et les données numériques accessibles. Ce processus de satisfaction de la qualité est évolutif et converge vers une situation satisfaisante pour l'utilisateur, alliant l'ontologie du problème et celle des données. La qualité externe est liée au calcul d'utilité qui provient de la comparaison entre la qualité attendue et la qualité interne des données disponibles. Nous utilisons deux matrices de qualité, représentant l'ontologie du problème et l'ontologie du produit, dans le même référentiel géosémantique. Il est possible, dans l'exemple présenté, d'évaluer la qualité externe de l'information géographique.

Ces travaux représentent une avancée dans ce domaine. Toutefois chacune de ces approches présente des limites lors de leur application :

– On suppose que les informations sur les données et leur qualité sont accessibles pour chaque couple à comparer. Pourtant cela est rarement possible, à moins de rester à un niveau de granularité général pour chaque élément de l'application, comme il est présenté dans l'exemple de l'application de Vienne où la qualité concerne les thèmes. Cependant, on pourrait penser à réaliser une classification fonctionnelle minimale des métadonnées permettant de mesurer, au minimum, le seuil critique d'évaluation de la qualité pour certains couples importants d'une application géospatiale ;

– Une autre limite importante de cette approche est que l'utilisateur doit être en mesure d'identifier les entités importantes de son application et de spécifier ses attentes de qualité, avec le seuil de tolérance pour les couples formés. Cela est contextuel à l'utilisateur, à l'application et à chaque couple. La question la plus difficile est d'expliciter chacune des deux ontologies et de les formaliser dans le même espace géosémantique. Pour aider à répondre à cette question, il serait envisageable d'établir une « bibliothèque de matrices de qualité », inspirées de cas déjà traités par des experts. Ces exemples de matrices guideraient l'utilisateur dans sa démarche de sélection des bonnes données en accord avec ses attentes de qualité. L'idée est d'identifier des couples (entité caractéristique, élément de qualité) qui sont présents dans plusieurs applications et de modéliser leur traitement ;

– Enfin, les trois dimensions « espace, thème, temps » sont essentielles pour analyser et comprendre les changements intervenant dans les problématiques géographiques dynamiques. Les modèles spatiaux négligent souvent la dimension temporelle : « en théorie, les ontologistes acceptent les objets, relations, états, et processus. En pratique, beaucoup de modèles sont représentés par des classes avec des attributs représentant un état avec certaines relations, laissant de côté les événements et processus » (traduction libre) [KUH 01]. Une ontologie intégrant la

dimension temporelle dans le domaine des systèmes d'information géographiques est une priorité (traduction libre) [FRA 03].

Les standards, comme l'ISO [ISO 02, ISO 03a, ISO 03b], n'échappent pas à cette remarque. Ils permettent de fournir une base pour mieux qualifier et quantifier la qualité externe : « ils aident les utilisateurs à évaluer dans quelle mesure des données correspondent à leur besoin » (traduction libre). Cependant, les standards proviennent de discussions et de négociations entre certaines institutions (agences cartographiques ou éditeurs de logiciels) et en ce sens fournissent une vision partielle de la réalité. Les standards de qualité de l'ISO couvrent en partie la description d'un état à un temps donné mais éludent la description de processus et d'événements, inhérente à certaines applications géospatiales. Aussi, pour étudier toutes les dimensions du problème de la qualité externe, il est important de décrire non seulement la qualité des besoins, des données, mais également des modèles décrivant les processus mis en jeu. Cette démarche est complexe mais importante à réaliser. Ce chapitre a tenté d'intégrer ces différentes dimensions.

15.5. Bibliographie

[AAL 98] AALDERS H., MORRISON J., « Spatial Data Quality for GIS » *Geographic Information Research: Trans-Atlantic Perspectives*, Craglia M., Onsrud H. (dir.), p. 463-475, Taylor & Francis, London, Bristol, 1998.

[AAL 99] AALDERS H., « The registration of Quality in a GIS » *Proceedings of the 1st International Symposium on Spatial Data Quality*, p. 23-32, Hong-Kong, 18-20 juillet 1999.

[AGU 98] AGUMYA A., HUNTER G.J., « Fitness for use: Reducing the Impact of Geographic Information Uncertainty » *Proceedings of the URISA 98 Conference*, p. 245-254, Charlotte, Etats-Unis, 18-22 juillet 1998.

[BED 95] BEDARD Y., VALLIERES D., Qualité des données à référence spatiale dans un contexte gouvernemental, Rapport technique, Université Laval, Canada, 1995.

[BRO 03] BRODEUR J., BEDARD Y., EDWARDS G., MOULIN B., « Revisiting the Concept of Geospatial Data Interoperability within the Scope of Human Communication Process », *Transactions in GIS*, vol. 7, nº 2, p. 243-265, 2003.

[CHR 83] CHRISMAN N.R., « The Role of Quality Information in the Long-Term Functioning of a Geographical Information System », *Proceedings of Auto Carto 6*, vol. 2, p. 303-321, Ottawa, 1983.

[COM 03] COMBER A., FISHER P., WADSWORTH R., « A Semantic Statistical Approach for Identifying Change From Ontologically Divers Land Cover Data », *Proceedings of the 6th AGILE Conference*, p. 123-131, Lyon, 24-26 avril 2003.

[DAS 03] DASSONVILLE L., VAUGLIN F., JAKOBSSON A., LUZET C., « Quality Management, Data Quality and Users, Metadata for Geographical Information », *Spatial Data Quality*, Shi W., Fisher P.F., Goodchild M.F. (dir.), p. 202-215, Taylor & Francis, 2003.

[DEB 01] DE BRUIN S., BREGT A., VAN DE VEN M., « Assessing fitness for use: the expected value of spatial data sets », *International Journal of Geographical Information Science*, vol. 15(5), p. 457-471, 2001.

[FRA 98] FRANK A.U., « Metamodels for data quality Description », *Data Quality in Geographic Information. From Error to Uncertainty*, Jeansoulin R., Goodchild M. (dir.), p. 15-29, Editions Hermès, Paris, 1998.

[FRA 03] FRANK A.U., A linguistically justified proposal for a spatio-temporal ontology, http://www.scs.leeds.ac.uk/brandon/cosit03ontology/position_papers/Frank.doc24-2003.

[FRA 04] FRANK A.U., GRUM E., VASSEUR B., « Procedure to Select the Best Dataset for a Task », *Proceedings of the Third International Conference on Geographic Information Science*, Egenhofer M.J., Miller H., Freksa C. (dir.), Univ. Maryland, Etats-Unis, 20-23 octobre 2004.

[GRU 93] GRUBER T.R., « A Translation Approach to Portable Ontology Specifications », *Knowledge Acquisition*, vol. 5, nº 2, p. 199-220, 1993.

[GRU 04] GRUM E., VASSEUR B., « How to select the Best Dataset for a Task? », *Proceedings of the International Symposium on Spatial Data Quality*, GeoInfo Series, vol. 28b, p. 197-206, Bruck an der Leitha, Autriche, 15-17 avril 2004.

[GUA 98] GUARINO N., « Formal Ontology and Information Systems », *Formal Ontology in Information Systems*, Guarino N. (dir.), IOS Press, p 3-15 ; Amsterdam, 1998.

[HUN 01] HUNTER G.J., « Keynote Address: Spatial Data Quality Revisited », *Proceedings of the 3rd Brazilian Geo-Information Workshop (Geo 2001),* Rio de Janeiro, 4-5 octobre 2001.

[HUN 02] HUNTER G.J., « Understanding Semantics and Ontologies: They're Quite Simple Really - If you know what I mean! », *Transactions in GIS*, vol. 6, nº 2, p. 83-87, 2002.

[ISO 00] ISO, ISO 9000 - Quality management systems, International Organization for Standardization (ISO), 2000.

[ISO 02] ISO/TC 211, 19113 Geographic information - Quality principles, International Organization for Standardization (ISO), 2002.

[ISO 03a] ISO/TC 211, 19114 Geographic information - Quality evaluation procedures, International Organization for Standardization (ISO), 2003.

[ISO 03b] ISO/TC 211, 19115 Geographic information - Metadata, International Organization for Standardization (ISO), 2003.

[JUR 74] JURAN J.M., GRYNA F.M.J., BINGHAM R.S., *Quality Control Handbook*, McGraw-Hill, New York, 1974.

[KOK 01] KOKLA M., KAVOURAS M., « Fusion of Top-level and geographical domain ontologies based on context formation and complementarity », *International Journal of Geographical Information Science*, vol. 15, nº7, p. 679-687, 2001.

[KUH 01] KUHN W., « Ontologies in support of activities in geographical space », *International Journal of Geographical Information Science*, vol. 15, nº7, p. 613-631, 2001.

[MAR 02] MARK D., EGENHOFER M., HIRTLE S., SMITH B., « Ontological foundations for geographic information science », *UCGIS Emerging Research Theme*, 2002. http://www.ucgis.org

[MOE 87] MOELLERING, H., « A draft Proposed Standard for Digital Cartographic Data », *National Committee for Digital Cartographic Standards*, American Congress on Surveying and Mapping Report #8, 1987.

[NOY 01] NOY N., MCGUINESS D., Ontology Development 101, « A guide to Creating your first Ontology », *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880*, Stanford University, Etats-Unis, 2001. http://protege.stanford.edu/publications/ontology_development/ontology101.html

[SHY 04] SHYLLON E.A., HUNTER G.J., « Fuzzy Optimization of Spatial Data Quality: Metadata for Dataset Searching », *Proceedings of the Third International Symposium on Spatial Data Quality*, GeoInfo Series, vol. 28a, p. 155-168, Bruck an der Leitha, Autriche, 15-17 avril 2004.

[SOW 98] SOWA J.F. *Knowledge Representation: Logical, Philosophical and Computational Foundations*, Brooks et Cole, Pacific Grove, Etats-Unis, 1998.

[UML 03] UML 2.0, Unified Modeling Language formalism, adopté par l'OMG http://www.omg.org/docs/ptc/03-09-15.pdf

[VAS 03] VASSEUR B., DEVILLERS R., JEANSOULIN R., « Ontological approach of the Fitness for Use of Geospatial datasets », *Proceedings of the 6th AGILE Conference*, p. 497-504, Lyon, 24-26 avril 2003.

[VAS 04] VASSEUR B., VAN DE VLAG D., STEIN A., JEANSOULIN R., DILO A., « Spatio-temporal Ontology for defining the quality of an application », *Proceedings of the International Symposium on Spatial Data Quality*, GeoInfo Series ,vol. 28b, p. 67-81, Bruck an der Leitha, Autriche, 15-17 avril 2004.

[VER 99] VEREGIN H., « Data quality parameters », *Geographical Information Systems*, Longley P.A., Goodchild M.F., Maguire D.J., Rhind D.W. (dir.), John Wiley & Sons, p. 177-189, 1999.