

Numéro d'identification:

UNIVERSITE DE MARNE-LA-VALLEE

THESE DE DOCTORAT

Spécialité: Sciences de l'Information Géographique

**Propagation des mises à jour dans les bases de données
géographiques multi-représentations par analyse des
changements géographiques**

par

Thierry Badard

Soutenue le 15 Décembre 2000, devant le jury composé de :

M. Yvan BEDARD, Professeur, Université Laval, Québec, Canada	Rapporteur
M. Jean-Paul CHEYLAN, Directeur de Recherche, CNRS	Directeur de Thèse
Mme. Geneviève JOMIER, Professeur, Université Paris IX Dauphine	Rapporteur
Mme. Sylvie LAMY, Ingénieur en Chef Géographe, DRAST - METL	Examineur
M. Daniel LAURENT, Professeur, Université de Marne-la-Vallée	Examineur
M. Hervé LE MEN, Habilité à Diriger des Recherches (HDR), IGN, Saint Mandé	Examineur
Mme. Thérèse LIBOUREL, Maître de Conférence, CNAM, LIRMM, Montpellier	Examineur

*A mes parents et à mon frère,
Et à toi, Emmanuel, que la mort a emporté si tôt ...
Mon rêve est devenu aujourd'hui réalité.*

Remerciements

Une thèse de doctorat n'est jamais une œuvre individuelle, même si le long et parfois difficile chemin qui mène au diplôme se parcourt pour l'essentiel seul, en faisant face à soi-même, à ses doutes et ses hésitations. Elle est le fruit de nombreuses expériences, rencontres, échanges, collaborations. Aussi voudrais-je ici exprimer toute ma gratitude à l'ensemble des personnes qui ont participé de près ou de loin à ce travail, en m'accordant, temps, écoute, conseils, expérience et parfois réconfort et m'auront permis à leur manière de progresser, avancer, m'enrichir, mûrir, m'améliorer, apprendre et me découvrir mieux encore moi-même.

Ainsi, je tiens ici à remercier, tout d'abord, le professeur Daniel Laurent, responsable de l'Institut Francilien d'Ingénierie des Services (IFIS) pour l'honneur qu'il a bien voulu me faire en acceptant de présider le jury de cette thèse de doctorat, ainsi que pour le grand intérêt qu'il a porté à mes travaux.

Toute ma gratitude et mes remerciements vont ensuite à mes deux rapporteurs :

Le professeur Geneviève Jomier, tout d'abord, du laboratoire LAMSADE de l'université Paris-Dauphine, pour l'honneur qu'elle a bien voulu me faire en donnant son avis d'expert en base de données sur ce travail de recherche, et ceci malgré le temps relativement restreint dont elle disposait du fait des délais de rédaction qui m'étaient imposés. Je tiens également à la remercier pour la collaboration active que nous avons eu durant cette thèse et que nous entretenons encore autour des problématiques de maintenance des bases de données géographiques multi-échelles.

Le professeur Yvan Bédard, ensuite, du Centre de Recherche en Géomatique de l'université Laval à Québec (Canada) qui m'a tant apporté et appris que ce soit dans le cadre du travail de recherche ou dans celui plus large de la vie. Avec toi Yvan, tout n'est que plaisir et ce concept, sans lequel rien ne vaut d'être fait, tu l'as fait devenir mien. Ce sont pour toutes ses choses qui dépassent largement tout ce que nos échanges techniques riches de toute ta maîtrise et de ton expérience ont pu déjà m'apporter, que je te serais éternellement reconnaissant.

Je remercie ici également mon directeur de thèse, monsieur Jean Paul Cheylan, directeur de recherche au CNRS, pour toute la « sensibilité » géographique qu'il a su me transmettre. Elle était en effet d'un intérêt primordial pour mener à bien le challenge que représentait cette thèse de doctorat et ceci malgré le nombre relativement restreint de fois où nous nous sommes réunis. Nos discussions m'ont apporté, même si tu as pu en douter, bien des éclairages sur comment orienter ou traiter tel point dur.

Mes remerciements s'adressent aussi à madame Thérèse Libourel, maître de conférence CNAM au LIRMM, qui a tout au long des trois années de cette recherche participé au comité de suivi de thèse et m'a apporté avec son extrême gentillesse et son franc parlé légendaires de nombreux conseils et critiques constructives, ainsi que toutes ses connaissances informatiques riches de plusieurs années à côtoyer les géographes. Un grand merci donc à toi, mon arlésienne préférée et maintenant que la soutenance est passée, tu peux à nouveau me tutoyer.

Mes remerciements les plus sincères vont également à monsieur Hervé Le men, directeur technique adjoint de l'Institut Géographique National pour tous les conseils judicieux et l'écoute qu'il m'a accordé lors de la rédaction dans un temps imposé très restreint, de ce mémoire de thèse sur travaux dont la structure m'était peu familière. De façon plus générale, je le remercie également pour l'intérêt qu'il a toujours porté à mes travaux de recherche et de développement.

Je tiens aussi à remercier madame Sylvie Lamy, ancienne responsable du laboratoire COGIT de l'IGN et maintenant chargé de mission à la Direction de la Recherche et des Affaires Scientifiques et Techniques (DRAST) du Ministère de l'Équipement, pour la chance qu'elle m'a donné de faire cette recherche au sein de son service et l'encadrement suivi qu'elle a réalisé tout au long de sa présence au COGIT. Je te remercie également pour la confiance et les responsabilités que tu m'as données pendant cette recherche, elles m'auront permis d'apprendre beaucoup même si mes journées me semblaient comporter plus de 24 heures parfois.

Je tiens aussi à remercier ici mesdemoiselles Anne Ruas, actuelle responsable du laboratoire COGIT, et Laure Dassonville pour les conseils, le soutien, les corrections, les encouragements et l'amitié qu'elles m'ont prodigués lors de la rédaction de ce mémoire et de sa relecture.

Mes remerciements vont également à l'ensemble de mes compagnons de route, qu'ils aient été thésards comme moi, stagiaires ou permanents, lors de cette thèse de doctorat. Un grand merci donc à Cécile Lemarié, Olivier Bucaille, Atef Bel Hadj Ali et Frédéric Hubert mes coturnes successifs au COGIT, Mohammed Ali Peerbocus et Laurent Spéry les braves de la recherche « made in France » en mise à jour des bases de données géographiques. Un grand merci également à tous mes amis : mon Cosmar, mon Nijo et sa Scalepa, Laurent C., Pierre-Jean et Isa, ML et sa MLinette, MP et sa mamour, Christophe et Ghislaine ...

Enfin, je tiens à remercier, même si cela vous semble le plus normal du monde, mes parents et mon frère Benoît pour l'écoute, la présence, le soutien, les encouragements et l'amour qu'ils m'ont apportés au jour le jour, que ce soit dans les moments d'euphorie et de joie que dans ceux de doutes et de remise en question.

Pour finir, un merci tout particulier à toi, Pascale. Grâce à toi, je me connais un peu plus moi-même et ceci malgré que cela se soit fait dans une très grande douleur et au pire moment ... Je te souhaite très sincèrement d'être, un jour, aussi heureuse que je le suis actuellement.

Résumé

De nos jours, les systèmes d'information géographique sont devenus de véritables outils d'aide à l'analyse et à la décision. De ce fait, un nombre croissant d'organisations s'en dotent et y incluent des informations spécifiques et utiles aux missions dont elles ont la charge. Cependant, la mise en place de tels systèmes est difficile et coûteuse. C'est pourquoi ces institutions font appel à des producteurs d'information géographique, afin d'acquérir des données de référence sur lesquelles elles vont appuyer leur système d'information. Ces organisations se placent ainsi, vis à vis du producteur, en tant qu'utilisateurs de données géographiques de référence.

Cependant, pour pouvoir répondre au mieux à leurs missions, ces institutions ont besoin de mises à jour de la part du producteur, afin de disposer d'un reflet de l'espace géographique qui soit le plus proche possible de la réalité du terrain, tout en conservant bien entendu leurs acquis. Il est à noter qu'un producteur peut avoir des préoccupations similaires, en voulant reporter les mises à jour de ces données de référence sur ses produits dérivés (bases à vocation cartographique ou à des échelles différentes), afin d'en diminuer les coûts de production. Mais à l'heure actuelle, cette propagation des évolutions spatio-temporelles du référentiel dans les bases de données géographiques utilisateur ou dérivées ne va pas sans poser des risques importants de perte d'information ou d'obtention d'états incohérents dans les bases.

Le but de cette thèse est ainsi d'apporter un ensemble de solutions aux problèmes entravant l'intégration des mises à jour dans de tels systèmes multi-représentations. Un mécanisme de mise à jour générique et automatique a été défini. Il a été mis en œuvre et validé sur des bases de données géographiques produites par l'Institut Géographique National. Ce travail a également permis la définition de modes de livraison de l'information d'évolution adaptés à la donnée géographique.

Mots-clefs

Bases de données géographiques, Mise à jour, Propagation, Multi-représentations, Livraison d'évolutions.

Abstract

Nowadays, Geographic Information Systems are considered to be truly analysis and decision-making tools. For that reason, a growing number of organisations invest in such systems and add specific information necessary to the tasks for which they have the responsibility. However, the implementation of such systems is difficult and relatively expensive. That is why these institutions purchase reference data sets from geographic information producers. From these they are able to develop their own information systems. Thus, these organisations are considered, by the producer, as users of reference geographic data sets.

To carry out their assignments, these institutions clearly need updates from the producer, in order to have the most faithful and realistic image of geographic reality. But these updates must allow for the preservation of consistency of all the knowledge added in the systems by users. Producers may have similar preoccupations if they want to propagate the updates from their reference data sets to their derived products (e.g. cartographic products or other databases with different scales) in order to reduce the production costs. However, at the present time, such a propagation of the effects of spatio-temporal evolutions from a reference database to a user's or a derived geographic database, may induce significant risks of information loss or leave the database in an inconsistent state.

This thesis tackles problems which hinder the integration of updates in such multi-representation systems. A generic and automatic updating mechanism has been defined. It has been implemented and validated on databases produced by the Institut Géographique National. Modes for the delivery of updating information have also been defined.

Keywords

Geographic databases, Update, Propagation, Multi-representations, Delivery of updating information.

Sommaire

CHAPITRE 1 INTRODUCTION	10
1.1 CONTEXTE ET OBJECTIFS	10
1.2 STRUCTURE DU DOCUMENT.....	11
CHAPITRE 2 VERS UN MÉCANISME DE MISE À JOUR	13
2.1 INTRODUCTION.....	13
2.2 LA NOTION DE MISE À JOUR	13
2.2.1 <i>Les évolutions terrain</i>	14
2.2.2 <i>Les corrections d'erreurs</i>	14
2.2.3 <i>Les changements de spécifications</i>	14
2.2.4 <i>Les enrichissements de la base</i>	15
2.3 LES PROBLÈMES ENTRAVANT LA MISE À JOUR DES SYSTÈMES D'INFORMATION GÉOGRAPHIQUE	15
2.3.1 <i>Les problèmes liés à l'utilisateur de données de référence</i>	15
2.3.1.1 Les modifications des données de référence par l'utilisateur.....	16
2.3.1.2 Une structuration de l'information ajoutée ou dérivée insuffisante	16
2.3.1.3 L'aspect multi-sources de l'information ajoutée.....	17
2.3.1.4 L'utilisateur qui devient producteur	18
2.3.2 <i>Les problèmes liés au producteur de données de référence</i>	19
2.3.2.1 Les modifications de spécifications	19
2.3.2.2 Les différences de fréquence de mise à jour	20
2.3.2.3 Le degré de liberté des mises à jour.....	20
2.3.3 <i>Les problèmes liés à la structuration des bases et à l'échange de l'information</i>	20
2.3.3.1 La structure même des bases de données géographiques.....	21
2.3.3.2 De l'interopérabilité des systèmes.....	21
2.4 LA MISE À JOUR DANS LES SYSTÈMES D'INFORMATION GÉOGRAPHIQUE	22
2.5 LE MÉCANISME DE MISE À JOUR DES BDG	25
2.5.1 <i>L'extraction des évolutions</i>	26
2.5.2 <i>L'extraction des relations de corrélation</i>	27
2.5.3 <i>L'intégration et la propagation des effets de la mise à jour</i>	30
2.6 CAS D'ÉTUDE	30
2.6.1 <i>Propagation des mises à jour de la BDCarto vers Route500</i>	31
2.6.2 <i>Propagation des mises à jour de la BDTopo vers la BDTopo25</i>	32
2.6.3 <i>Présentation de GéO₂</i>	34
CHAPITRE 3 EXTRACTION DES MISES À JOUR ET DIFFUSION DE L'INFORMATION D'ÉVOLUTION	36
3.1 INTRODUCTION.....	36
3.2 CONTRIBUTIONS.....	36
3.2.1 <i>Mécanisme d'extraction des mises à jour et première proposition pour leur échange (lots différentiels)</i>	36
3.2.2 <i>Echange de l'information d'évolution via l'utilisation de XML (lots d'évolution)</i>	37
3.3 COMPLÉMENTS.....	72
3.3.1 <i>Retour sur l'appariement de données géographiques</i>	72
3.3.1.1 Justification du choix de méthodes d'appariement	72
3.3.1.2 Fonctionnement détaillé du processus générique d'appariement.....	72
3.3.2 <i>Retour sur le mécanisme d'extraction des mises à jour</i>	76
3.3.2.1 Problèmes de robustesse du processus d'appariement.....	76
3.3.2.2 Notion de « traçabilité » de l'information d'évolution	77
3.4 RÉSULTATS DES CAS D'ÉTUDE.....	77
3.4.1 <i>Extraction des mises à jour de la BDCarto</i>	78
3.4.2 <i>Extraction des mises à jour de la BDTopo</i>	79
CHAPITRE 4 INTÉGRATION ET PROPAGATION DES MISES À JOUR	82
4.1 INTRODUCTION.....	82
4.2 CONTRIBUTIONS.....	82
4.3 COMPLÉMENTS.....	95

4.3.1	<i>Modèle de gestion de l'intégration et de la propagation</i>	95
4.3.2	<i>Précisions sur la gestion des conflits de mise à jour</i>	98
4.4	RÉSULTATS DES CAS D'ÉTUDE	98
4.4.1	<i>Propagation des mises à jour de la BDCarto vers Route500</i>	99
4.4.2	<i>Propagation des mises à jour de la BDTopo vers BDTopo25</i>	101
	CHAPITRE 5 CONCLUSION ET PERSPECTIVES	104
5.1	CONCLUSION	104
5.2	PERSPECTIVES	106
	BIBLIOGRAPHIE	108
	PUBLICATIONS	114

Table des figures

<i>Figure 2.1 - Illustration du manque de relations explicites entre couches d'information dans les bases de données géographiques</i>	<i>17</i>
<i>Figure 2.2 - Exemple de données de degrés de qualité géométrique différents</i>	<i>18</i>
<i>Figure 2.3 - Structure du mécanisme de mise à jour des bases de données géographiques</i>	<i>26</i>
<i>Figure 2.4 - Exemple de relations de partage de géométrie retrouvées entre les thèmes administratif et hydrographie de la BDCarto®.....</i>	<i>29</i>
<i>Figure 2.5 - Différences de contenu des thèmes routiers BDCarto et Route500</i>	<i>32</i>
<i>Figure 2.6 - Différences de représentation entre la BDTopo (à gauche) et la carte au 1 : 25 000 (TOP25, ©IGN, à droite).....</i>	<i>33</i>
<i>Figure 3.1 - Enchaînement des phases du processus générique d'appariement.....</i>	<i>73</i>
<i>Figure 3.2 - Calque différentiel (objets contenus dans BDCARTOV2).....</i>	<i>78</i>
<i>Figure 3.3 - Calque différentiel (objets contenus dans BDTPOV2).....</i>	<i>80</i>
<i>Figure 4.1 - Modèle de gestion de l'intégration et de la propagation des mises à jour.....</i>	<i>95</i>
<i>Figure 4.2 - Carte des conflits d'intégration des mises à jour BDCarto dans Route500.....</i>	<i>99</i>
<i>Figure 4.3 - Interface de résolution des conflits de mise à jour</i>	<i>100</i>
<i>Figure 4.4 - Résultats du test de propagation des mises à jour de la BDTopo sur la BDTopo25 (extrait 1).....</i>	<i>101</i>
<i>Figure 4.5 - Résultats du test de propagation des mises à jour de la BDTopo sur la BDTopo25 (extrait 2).....</i>	<i>102</i>
<i>Figure 4.6 - Illustration des décalages de géométrie entre versions à jour de la BDTopo25.....</i>	<i>103</i>

Liste des tableaux

<i>Tableau 2.1 - Différences de contenu en nombre d'objets entre la BDCarto et Route500</i>	32
<i>Tableau 2.2 - Différences de contenu en nombre d'objets entre la BDTopo et la BDTopo25</i>	34
<i>Tableau 3.1 - Résultats détaillés de l'extraction des mises à jour entre les deux versions de la BDCarto</i>	79
<i>Tableau 3.2 - Résultats détaillés de l'extraction des mises à jour entre les deux versions de la BDTopo d'Angers</i>	81

Chapitre 1

Introduction

1.1 Contexte et objectifs

De nos jours, les systèmes d'information géographique (SIG) sont devenus de véritables outils d'aide à l'analyse, à la compréhension, à la gestion du territoire ainsi qu'à la planification et à la décision. De ce fait, un nombre croissant d'organisations s'en dotent et y incluent des informations spécifiques, qui peuvent être de natures très diverses, utiles aux missions dont elles ont la charge. Cependant, la mise en place de tels systèmes est difficile et relativement coûteuse. C'est pourquoi ces institutions font appel à des producteurs d'information géographique, tel que l'Institut Géographique National (IGN) en France. Elles peuvent donc acquérir des données de référence sur lesquelles elles vont pouvoir appuyer leur système d'information. Ces organisations se placent ainsi, vis à vis du producteur, en tant qu'utilisateurs de données géographiques de référence (appelées aussi référentiels).

Cependant, pour pouvoir répondre au mieux aux missions dont ils ont la charge, ces utilisateurs ont besoin de disposer d'un reflet de l'espace géographique qui soit le plus proche possible de la réalité du terrain. Ils nécessitent donc, de la part du producteur, la livraison de mises à jour concernant les données de référence qu'ils puissent facilement intégrer au sein de leurs systèmes d'information et qui ne viennent pas remettre en cause la connaissance qu'ils y ont déjà capitalisée.

Un producteur de données géographiques peut avoir des préoccupations similaires. En effet, à partir des bases de référence qu'il conçoit, maintient et entretient, celui-ci dérive un certain nombre d'autres produits numériques à vocations diverses (production de cartes papier, de bases de données présentant des niveaux d'échelle ou de représentation différents, etc.). Afin de minimiser les coûts de production et de maintenance de tels produits, celui-ci désirera reporter les évolutions de ces données de référence sur ses produits dérivés.

Cependant, à l'heure actuelle, cette intégration et cette propagation des mises à jour d'un référentiel dans des bases de données géographiques (BDG) utilisateur ou dérivées ne va pas sans poser des risques importants de perte d'information ou d'obtention d'états incohérents dans celles-ci. Ces problèmes sont principalement liés à l'utilisation même qui est faite des données de référence et au manque de modes de livraison adaptés permettant de mettre à disposition l'information de mise à jour.

L'objectif de cette thèse est donc d'apporter un ensemble de solutions permettant de **réaliser une intégration des mises à jour et une propagation des effets de celles-**

ci dans les bases de données géographiques utilisateur ou dérivées, qui soit la plus automatique possible et qui n'entraîne pas de dégradations de la cohérence et des acquis dans de tels systèmes, la diversité des processus et des traitements impliqués dans leur conception permettant de qualifier ces derniers de systèmes multi-représentations [Brugger et. al., 1989]. Un mécanisme de mise à jour de la donnée géographique vecteur répondant à cet objectif et s'étendant de la détection des évolutions au sein des bases de données géographiques de référence, jusqu'à l'intégration et la propagation des mises à jour dans les systèmes utilisateur ou dérivés, a été ainsi mis au point et est présenté dans ce mémoire. Les résultats de différents cas d'étude basés sur des données produites par l'IGN et ayant permis de tester et valider ce mécanisme, sont également fournis et détaillés. La structure de ce mémoire de thèse sur travaux (i.e., incluant un certain nombre d'articles publiés dans des revues ou actes de conférences internationales) est donnée dans la section suivante.

1.2 Structure du document

Le chapitre 2 s'attache tout d'abord à définir ce que recouvre le terme de mise à jour dans les bases de données géographiques afin de bien comprendre les difficultés que soulève cette opération. L'ensemble des problèmes entravant l'intégration et la propagation des mises à jour dans les bases de données utilisateur ou dérivées, est alors détaillé. Les éléments qui permettraient d'aider à leur résolution sont également mis en exergue. S'appuyant sur cette taxonomie, un état de l'art des solutions existantes est ensuite dressé et montre en quoi ces solutions ne répondent que très partiellement à la question. Au vu des problèmes identifiés et des apports fournis par les solutions existantes, la structure et le fonctionnement du mécanisme de mise à jour que nous proposons, sont alors décrits. La description succincte des techniques qu'il met en œuvre, depuis l'extraction des évolutions jusqu'à la propagation des mises à jour dans les bases de données utilisateur ou dérivée, permettra de montrer que l'ensemble des problèmes que pose l'opération de mise à jour est adressé. Les cas d'étude ayant permis de tester et valider ce mécanisme sont enfin présentés.

Le chapitre 3 présente de façon détaillée le processus d'extraction des évolutions entre deux versions d'une même base de données géographiques que nous avons mis au point et sur lequel s'appuie le mécanisme de mise à jour proposé. Ce processus constitue un préalable indispensable à toute propagation des mises à jour dans les systèmes. L'information qu'il délivre a permis la définition de modes de livraison, permettant de mettre à disposition l'information d'évolution pour les bases de données géographiques. La description détaillée de ceux-ci et de leurs avantages et inconvénients, sera fournie également dans ce chapitre. Un certain nombre de compléments et de questions connexes que met en exergue le processus d'extraction sera fourni au lecteur afin d'en compléter la compréhension des forces et faiblesses. Enfin, les résultats des extractions de mises à jour opérées sur les différents cas d'étude entrepris, seront donnés afin d'illustrer son fonctionnement.

Le chapitre 4 présente le processus d'intégration et de propagation des mises à jour dans les bases de données géographiques qui est au cœur du mécanisme proposé dans ce mémoire. L'ensemble des détails concernant sa définition et son fonctionnement (notamment les procédés mis en place pour permettre la conservation des acquis et de la cohérence dans les systèmes), sera donné dans ce chapitre. Les résultats obtenus sur les cas d'étude y seront également fournis et discutés.

Enfin, le chapitre 5 fournira une synthèse des réflexions et propositions menées dans le cadre de ce travail de thèse et fournira un certain nombre de perspectives visant non seulement à améliorer le fonctionnement du mécanisme proposé, mais aussi d'adresser d'autres problématiques liées à la mise à jour des bases de données géographiques.

Chapitre 2

Vers un mécanisme de mise à jour

2.1 Introduction

Ce chapitre s'attache tout d'abord à définir ce que recouvre le terme de mise à jour dans les bases de données géographiques (section 2.2) afin de bien comprendre les difficultés que soulève cette opération. L'ensemble des problèmes entravant l'intégration et la propagation des mises à jour dans les bases de données utilisateur ou dérivées, est alors détaillé (section 2.3). Les éléments qui permettraient d'aider à leur résolution sont également mis en exergue au sein de cette même section. S'appuyant sur cette taxonomie, un état de l'art des solutions existantes est ensuite dressé et montre en quoi ces solutions ne répondent que très partiellement à la question (section 2.4). Au vu des problèmes qui ont été identifiés et des apports fournis par les solutions existantes, la structure et le fonctionnement du mécanisme de mise à jour que nous proposons, sont alors décrits (section 2.5). La description succincte des techniques qu'il met en œuvre, depuis l'extraction des évolutions jusqu'à la propagation des mises à jour dans les bases de données utilisateur ou dérivée, permettra de montrer que l'ensemble des problèmes que pose l'opération de mise à jour est adressé par celui-ci. Les cas d'étude ayant permis de tester et valider ce mécanisme sont enfin présentés (section 2.6).

2.2 La notion de mise à jour

Souvent définie dans la littérature [Dell'Erba et Libourel, 1997 ; Spéry, 1999, p. 77-80], sous la forme d'évolutions de schéma et/ou de données, nous déciderons ici de définir la mise à jour dans le cadre des bases de données géographiques sous un angle permettant de remonter à la nature de la modification à apporter. En effet, même si cela se traduira, au niveau de la mise en œuvre dans les systèmes, par des modifications de structuration de l'information ou de l'information elle-même, la nature des évolutions possède une réalité physique et une sémantique forte, liée à celle de l'information modélisée (à savoir des phénomènes du monde réel), qui induira des modifications dans les bases de données qui n'ont pas la même signification, et donc pas forcément les mêmes effets. Il nous semble donc important de s'attacher aux causes des évolutions, afin de ne pas en perdre le sens, les répercussions dans une base de données utilisateur ou dérivée en dépendant fortement.

Ainsi, le terme de mise à jour de données géographiques, afin de traiter le problème dans son ensemble, se doit de recouvrir les notions d'évolution terrain, de correction d'erreurs, de changement de spécifications et d'enrichissement éventuel de la base de données. Ces différents types d'évolution sont détaillés les sections suivantes.

2.2.1 Les évolutions terrain

Les évolutions terrain sont ce que nous pourrions appeler les « véritables évolutions géographiques ». Elles traduisent, en effet, toutes les évolutions subies par les entités du monde réel modélisées dans les bases de données géographiques¹. Elles rendent donc compte de la dynamique spatiale, c'est-à-dire des changements observables et observés, intervenus dans le paysage, qu'ils soient naturels ou « provoqués » (i.e. dus à l'action de l'homme ou causés par d'autres phénomènes naturels).

Ce type d'évolution se traduira dans les systèmes uniquement par des créations, suppressions, modifications d'objets ou d'informations stockés dans les bases de données.

2.2.2 Les corrections d'erreurs

Elles correspondent à des corrections, effectuées par le producteur ou par l'utilisateur, du jeu de données fourni lors d'une précédente livraison, et ceci même si aucun changement n'est intervenu sur le terrain. Ces erreurs sont, par exemple, l'oubli d'un objet géographique lors de la saisie de la base, le mauvais renseignement d'une valeur d'attribut ou une mauvaise localisation d'un objet. Les corrections d'erreurs viennent donc modifier l'état de la modélisation du monde réel géré dans les bases de données géographiques, afin de refléter au mieux la réalité du terrain.

Les corrections d'erreurs, comme les évolutions terrain, concerneront uniquement les objets et l'information contenus dans les bases de données géographiques.

2.2.3 Les changements de spécifications

Il s'agit essentiellement de modifications apportées par le producteur de données afin de simplifier ou d'enrichir la perception du monde réel modélisé dans la base de données, et ainsi en modifier la qualité et/ou la richesse. Ils se décomposent en des changements des spécifications de saisie (ex. : modification de la longueur minimale de saisie d'un arc géométrique pour les objets géographiques linéaires) et de contenu (ex. : regroupement de deux classes sémantiquement proches dans une seule et même classe d'objets). Les changements de spécification affecteront donc les bases de données non seulement au niveau de leur schéma mais aussi au niveau de leur contenu, c'est-à-dire les données stockées dans celles-ci.

¹ Une base de données géographiques n'est en effet qu'une abstraction du monde réel. Les entités modélisées dans celle-ci, ne reflètent les phénomènes géographiques observables que conformément au filtre des spécifications établies par le producteur. On qualifie souvent de « terrain nominal », le résultat de cette abstraction [David et Fasquel, 1997].

2.2.4 Les enrichissements de la base

Sous ce terme, nous entendons l'intégration de thèmes non contenus dans la livraison précédente réalisée par le producteur des données géographiques de référence, mais aussi l'ajout d'information provenant de bases de données exogènes qu'un utilisateur a utilisées pour enrichir son système d'information. Par exemple, un utilisateur en charge de la gestion du réseau routier, peut demander à un producteur la livraison d'un thème relatif à l'hydrographie, non livré jusqu'alors, car ses attributions ont évolué. Il peut, de la même façon, utiliser des sources de données numériques produites par un autre de ses services, afin de mener à bien une étude particulière.

Ces enrichissements de la base peuvent également être le fait d'une extension du domaine d'étude. Le producteur augmentera alors son emprise d'extraction dans sa base de données, afin de pourvoir à la nouvelle requête de son client. Il ne s'agit donc pas d'évolutions terrain, ni de changements dans les spécifications du produit à proprement parlé.

Ces évolutions auront des répercussions sur la structuration de l'information contenue dans la base (extension du schéma, ajout de relations, ...) mais également sur l'information elle-même (ajout de nouveaux objets et de nouveaux liens entre les instances).

La nature des mises à jour dans le contexte des bases de données géographiques étant maintenant clairement définie, il convient de s'attacher à bien isoler les problèmes entravant son intégration dans les systèmes, l'établissement d'un mécanisme de mise à jour générique de la donnée géographique ne pouvant se faire qu'à ce prix. Ceci fait l'objet de la section suivante. La taxonomie des problèmes liés à l'opération de mise à jour qui y est établie s'appuie sur [Badard, 1998a ; Badard, 1998b].

2.3 Les problèmes entravant la mise à jour des systèmes d'information géographique

Ces problèmes peuvent se classer en trois grandes catégories que nous avons là aussi volontairement formulées en nous attachant aux causes malgré le caractère quelque peu redondant de la forme des problèmes qui se posent, car la résolution de ces derniers en dépend fortement.

2.3.1 Les problèmes liés à l'utilisateur de données de référence

L'ensemble de ces problèmes résulte de l'utilisation qui est faite de la donnée géographique. Ils proviennent de la modification des données de référence par les utilisateurs, d'une structuration insuffisante de l'information qu'ils ajoutent ou dérivent dans leurs systèmes, de la qualité relative des données qu'ils importent ou produisent et du retour qu'ils peuvent parfois inférer en terme de mise à jour, sur le producteur de données géographiques.

2.3.1.1 Les modifications des données de référence par l'utilisateur

Lorsqu'un utilisateur dispose d'une forte maîtrise ou est en charge de la gestion des entités du monde réel modélisé dans sa base de données, celui-ci peut être amené à modifier le référentiel. Or, ces modifications peuvent souvent devenir des difficultés supplémentaires pour l'intégration de mises à jour. En effet, il se peut par exemple que la version initiale de l'objet ne soit pas conservée car remplacée par la version modifiée dans le référentiel lui-même ou que si un identifiant (ou référence) externe existe sur les objets, la filiation entre les différentes versions des objets de la base ne soit pas ou mal assurée. La mise à jour peut alors entraîner irrémédiablement de la perte d'information ou l'obtention d'un état incohérent de la base, car elle aura tendance à remplacer les objets modifiés par l'utilisateur au profit de ceux contenus dans le nouvel état du référentiel. Il faut donc pouvoir se doter de mécanismes de mise à jour capables de gérer ce genre de conflits.

D'autre part, le référentiel peut avoir été complètement transformé par l'utilisateur, au point que le schéma et les données de la base de référence n'ont plus d'existence réelle dans son système, tout au moins dans l'état dans lequel le producteur l'a livré de façon initiale. C'est par exemple le cas, dans les systèmes de production cartographique où pour répondre au mieux au but visé, à savoir l'édition de carte papier, le schéma de la base de référence, ainsi que les objets qu'elle contient subissent de nombreux traitements (simplification, généralisation [Brassel et Weibel, 1988 ; Ruas, 1999], symbolisation, etc.). Or, tous ces processus ne gardent généralement pas trace des liens qu'ils ont inférés entre le résultat et les objets ou classes de référence. Cela impose donc soit de retrouver ces liens pour pouvoir intégrer les mises à jour, soit de dériver à nouveau une nouvelle actualité de la base de référence. Ceci pose donc le problème du manque de structuration des systèmes utilisateur ou dérivés.

2.3.1.2 Une structuration de l'information ajoutée ou dérivée insuffisante

Dans la plupart des systèmes d'information géographique mis en place par les utilisateurs, très peu de liens explicites entre les données de référence et celles que rajoute ou dérive l'utilisateur (ou même entre les couches d'information additionnelles) sont établis. Tout se passe un peu comme si on empilait des calques les uns par dessus les autres. Ceci est principalement lié aux latitudes offertes par les logiciels de SIG du marché, qui ne s'appuient pas, pour la plupart, sur de véritables systèmes de gestion de base de données (SGBD) et n'offrent donc pas à l'utilisateur les moyens de structurer fortement son système. Il est, en outre, à noter que cette possibilité de structuration est souvent ignorée par ceux-ci, soit par méconnaissance des potentialités du logiciel, soit par la non prise en compte de la notion de mise à jour lors de la phase de conception du système.

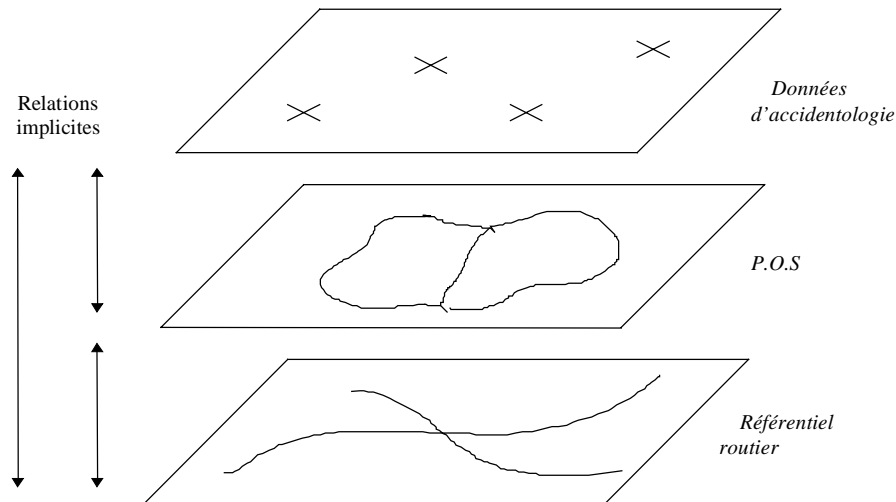


Figure 2.1 - Illustration du manque de relations explicites entre couches d'information dans les bases de données géographiques

La Figure 2.1 fournit un exemple d'un tel système. Les données d'accidentologie, s'appuyant pourtant sur le tracé des routes ne sont, en aucun cas, explicitement liées aux données routières contenues dans le référentiel. Il en va de même pour les informations localisées concernant le Plan d'Occupation des Sols (POS), qui pourtant possède des parties de géométrie commune avec le réseau routier.

Ce manque de structuration rend très difficile la propagation des effets de la mise à jour du référentiel dans la base de données géographiques de l'utilisateur. La mise à jour peut, en effet, conduire irrémédiablement à de la perte d'information ou à l'obtention d'un état incohérent (d'ordre logique) de la base. Dans l'exemple illustré précédemment, toute modification de la géométrie, intégrée au référentiel et portant sur une partie du réseau routier comportant des données d'accidentologie n'entraînera pas la mise à jour de ces dernières. On se retrouvera alors, après intégration des mises à jour dans le référentiel, avec des accidents qui ne seront plus localisés sur un tronçon de route. Cette conséquence est une grave violation de la cohérence de la base de données géographiques utilisateur.

La notion de préservation de la cohérence dans les bases de données géographiques, du fait de la nature de l'information manipulée, est ainsi bien plus large que dans les bases de données dites classiques. En effet, du fait de la composante géométrique dont disposent les objets, celle-ci doit englober le respect des contraintes d'intégrité spatiale [Scholl et al., 1996, pp. 25-26 ; Longley et al., 1999, pp. 85-87, Laurini et Thompson, 1992, pp. 175-214]. Ainsi, il apparaît que lors de la mise à jour des systèmes utilisateur, si l'on veut préserver la cohérence d'ensemble, toutes les relations implicites (spatiales ou non) entre les données stockées doivent être, au préalable, retrouvées et établies.

2.3.1.3 L'aspect multi-sources de l'information ajoutée

Les différents degrés de qualité, pris dans un sens très large, des données ayant permis l'enrichissement du référentiel pour réaliser la base utilisateur ne va pas non plus sans poser de problèmes. En effet, on peut trouver dans la base utilisateur des données, qui lui sont propres, d'échelles, de précisions géométriques différentes et de

spécifications variables qui vont accroître les problèmes de préservation de la cohérence lors de la propagation des mises à jour du référentiel sur les données utilisateurs.

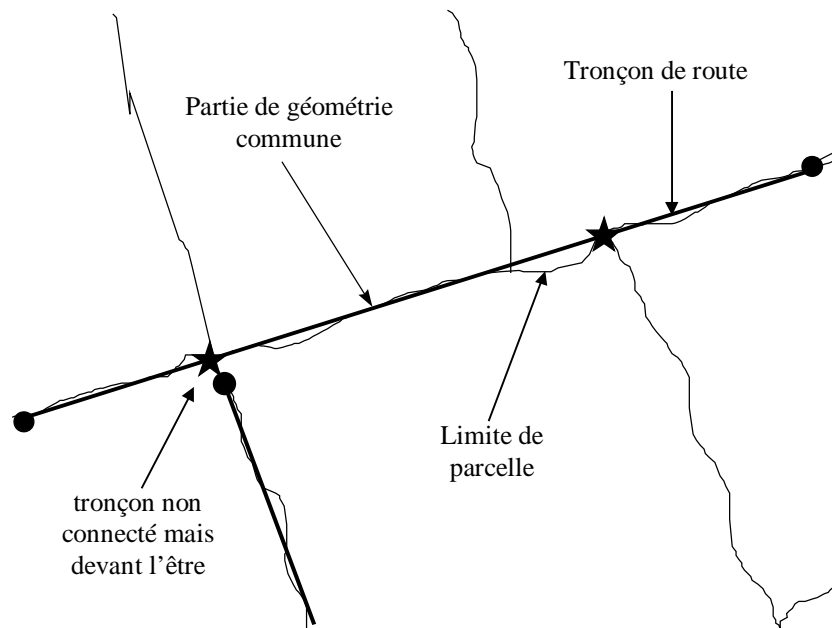


Figure 2.2 - Exemple de données de degrés de qualité géométrique différents

La Figure 2.2 illustre le cas de données routières et parcellaires, saisies ou importées par l'utilisateur dans des couches différentes et qui s'avèrent non cohérentes entre elles. La topologie des parties de réseau routier n'est pas assurée et les limites de parcelles ne partagent pas la géométrie des portions de routes qui pourtant les délimitent.

Les logiciels de SIG du marché sont pourtant, pour la plupart, dotés d'outils de récupération de la topologie, mais sont soit non utilisés, soit peu robustes ou inopérants, les écarts de géométrie étant dans certains cas extrêmes.

Les processus d'établissement de relations entre objets géographiques, mis en exergue dans la section précédente, et qui permettraient de propager les effets de la mise à jour du référentiel sur les données utilisateur ou dérivées, sans perte d'information ou obtention d'états incohérents, doivent donc tenir compte de ces écarts de qualité entre les données contenues dans les systèmes, en permettant de disposer de certains degrés de tolérance dans la recherche des objets à lier.

2.3.1.4 L'utilisateur qui devient producteur

La frontière entre producteurs et utilisateurs de données de référence peut devenir moins franche, car ces derniers peuvent se placer à leur tour en tant que producteur de données en communiquant tout ou partie de l'information qu'ils possèdent à d'autres institutions ou voire instaurer un retour vers le ou les producteurs auprès desquels ils ont acquis leurs données de référence. Il se pose alors des problèmes particuliers comme la récupération de la géométrie d'un objet (après prise en compte par le producteur dans sa base de données et re-livraison aux utilisateurs) tout en conservant l'information sémantique spécifique de l'objet utilisateur. Un processus de mise à jour devra donc être

capable de gérer ces conflits particuliers, en permettant une certaine réactivité sur les mises à jour à intégrer ou non dans les systèmes.

D'autres difficultés relevant du producteur de données de référence, viennent s'ajouter aux problèmes liés à l'utilisation de celles-ci. Elles sont présentées dans la section suivante.

2.3.2 Les problèmes liés au producteur de données de référence

Ces problèmes proviennent des modifications apportées par le producteur aux spécifications de ses bases de données de référence, aux différences de rythme de mise à jour entre les différentes données qu'il produit, et à la façon même dont il applique ses changements au sein de ses bases.

2.3.2.1 Les modifications de spécifications

Ces modifications peuvent concerner le référentiel lui-même ou la forme des livraisons des mises à jour. Dans le premier cas, il s'agit soit de modifications des spécifications de contenu (par exemple, les autoroutes et les routes ne forment plus qu'une seule et même classe) ou de saisie (par exemple, l'augmentation de la longueur minimale d'un arc, ce qui implique la seule modification des instances). Bien que la plupart du temps documentés, ces changements (ajout, suppression, scission, agrégation de structures et/ou d'information) affectent de façon lourde la structuration de l'information contenue dans la base utilisateur et vont l'obliger à la réorganiser dans sa plus grande partie en établissant de nouveaux liens non seulement entre les classes de son schéma mais aussi entre les objets eux-mêmes (à schéma constant dans le cas de changements de spécifications de saisie). Un processus de mise à jour prenant en compte ce problème se devra donc d'assister l'utilisateur dans l'intégration de ces changements au sein de sa base.

Le changement des spécifications de la forme de livraison des mises à jour est un problème en devenir, car actuellement, aucun mode de livraison des évolutions n'est communément mis en œuvre pour la donnée géographique. En effet, cette information de mise à jour est le plus souvent délivrée sous la forme de la base entière à jour, et ceci même si le taux d'évolution d'une base de données géographiques est communément estimé à 10% par an [Raynal, 1996]. Les mises à jour se retrouvent ainsi disséminées dans un volume important de données n'ayant pas évolué. Un utilisateur désireux d'intégrer les modifications dans son système d'information a alors le choix entre deux alternatives. Il peut soit remplacer entièrement l'ancienne version de son jeu de données de référence par la nouvelle, soit extraire par lui-même les évolutions subies par les entités géographiques modélisées dans la base de données, par comparaison des deux états de celle-ci. Cependant, la première solution peut entraîner, comme nous l'avons vu (cf. section 2.3.1), des pertes importantes d'information ou l'obtention d'états incohérents de la base utilisateur ou dérivée après intégration des mises à jour. La deuxième solution, peut quant à elle s'avérer difficile à mettre en œuvre, voire impossible, si un système fiable et stable d'identification des objets ou de gestion de l'historique n'est pas implanté dans les bases de données mises en jeu.

Ce manque de modes de livraison de l'information d'évolution pose donc le problème de l'accès et de la mise à disposition de cette information pour la donnée

géographique, derrière lesquels se cachent en fait la difficulté d'expression de celle-ci. En effet, s'il est relativement aisé de traduire dans le cadre des bases de données dites « classiques », les modifications attributaires subies par les objets, il n'en va pas de même pour l'information géométrique stockée dans les bases de données géographiques. Cela tient donc en grande partie à la spécificité de l'information manipulée. Afin de permettre l'accès et la mise à disposition des évolutions dans les bases de données géographiques, des mécanismes d'extraction et des modes de livraison de cette information devront être définis.

2.3.2.2 Les différences de fréquence de mise à jour

Il est compréhensible que sur l'étendue du domaine géré par un utilisateur, les thèmes et certaines zones évoluent à des vitesses différentes, ce qui va induire des rythmes de mise à jour différents entre les données, ce qui peut poser des problèmes complexes lorsque ces informations sont en relation avec d'autres données de la base (cf. section 2.3.1.2). On risque notamment d'avoir des états transitoires incohérents ou qui ne reflètent pas la réalité. Un processus de mise à jour de la donnée géographique devra donc être capable de repérer ces incohérences temporelles et permettre à l'utilisateur de pouvoir les résoudre, afin de préserver la cohérence de son système.

2.3.2.3 Le degré de liberté des mises à jour

Dans le cadre par exemple du Cadastre, les évolutions sont fortement contraintes car réglementées par le législateur [Spéry, 1999, pp. 36-41]. Leur nombre est donc limité et leurs natures connues à l'avance. Le processus de propagation des mises à jour sur les données utilisateur sera donc a priori plus facilement définissable et maîtrisable. Dans le cadre de bases de données géographiques où les transformations sont plus libres, la propagation sera alors plus difficile à appréhender, car on ne dispose que de peu d'informations sur la nature des évolutions subies par les entités modélisées (problème de disponibilité et d'accès à l'information de mise à jour, cf. section 2.3.2.1) et leurs types sont quasi infinis (problème d'expression de l'information d'évolution, cf. section 2.3.2.1). La diversité des thèmes et des transformations implique donc une forte maîtrise et connaissance, de la part de l'utilisateur, des phénomènes modélisés dans les bases et des processus de transformation. Un mécanisme de mise à jour se devra donc de capitaliser cette connaissance et de permettre le déclenchement de traitements appropriés (dérivés de ceux ayant conduit à la conception des systèmes), afin de mener à bien sa tâche, sans perte d'information et sans dégradation de la cohérence.

Maintenant définis les problèmes liés au producteur de données de référence, d'autres difficultés liées à l'échange et à la structuration de l'information géographique aux sein des systèmes, sont présentées dans la section suivante.

2.3.3 Les problèmes liés à la structuration des bases et à l'échange de l'information

Ces problèmes proviennent de l'absence dans certaines bases de données de systèmes d'identification stables et fiables, des limites des formats d'échange actuels à pouvoir délivrer l'information d'évolution, et des niveaux différents de structuration de l'information géographique que peuvent gérer les SIG.

2.3.3.1 La structure même des bases de données géographiques

L'inexistence de systèmes fiables et stables d'identification des objets dans certaines bases de données géographiques ne va pas sans poser, nous l'avons vu (cf. section 2.3.2.1), des problèmes d'accès à l'information d'évolution. Néanmoins, ce manque peut se justifier. En effet, du fait des relations topologiques et de composition qui peuvent exister entre les entités stockées dans une base de données géographiques, toute modification d'un objet a tendance à se propager sur les objets avec lesquels il est lié. La mise en place de tels systèmes impose donc une gestion très stricte et des efforts de maintenance considérables de la part du producteur. C'est pourquoi, nombres de bases de données géographiques n'en disposent pas. Un mécanisme de mises à jour devra donc pouvoir s'absoudre de l'existence de tels systèmes.

2.3.3.2 De l'interopérabilité des systèmes

Comme nous l'avons vu dans la section 2.3.2.1, la plupart des formats d'échange actuels de l'information géographique ne permettent pas de délivrer l'information de mise à jour. Ils n'ont en effet pas été conçus pour cela : leur vocation première était de permettre l'échange de jeu de données statiques, et non pas l'échange d'information dynamique, comme peut l'être l'information d'évolution, c'est-à-dire permettant la description des modifications subies par les objets. Si certains formats d'échange standardisés de la donnée géographique peuvent le permettre, cette livraison ne se fait qu'au prix de la définition d'un contenu spécifique et dédié, jouant avec les limites du format. C'est le cas, par exemple, d'EDIGéO [AFNOR, 1992] qui reporte cette gestion à un niveau attributaire, en livrant des géométries fictives (i.e., qui n'ont aucune existence physique dans la réalité) et dans un mode non topologique, ce qui représente un danger pour un utilisateur non averti qui intégrerait brutalement ces mises à jour dans sa base.

Le niveau de structuration de l'information gérée par les systèmes peut également être un handicap. La gestion de la structure de l'information est en effet réalisée à des degrés très divers d'un système à un autre. Certains systèmes par exemple ne gèrent pas la topologie ou le partage de géométrie, ou ne permettent pas l'accès aux primitives géométriques, ou encore la manipulation d'objets complexes (i.e., composés d'objets simples). On se heurte donc ici à des problèmes liés à l'interopérabilité des systèmes d'information géographiques. En effet, le simple fait qu'un système récepteur de l'échange de mise à jour gère un niveau de structuration de l'information différent de celui de l'échange, rendra toute mise à jour de celui-ci impossible.

De la même façon, la livraison d'information d'évolution peut se heurter à des problèmes d'interopérabilité sémantique [Kuhn, 1994 ; Bishr, 1998]. En effet, un utilisateur qui aura modifié la structuration du référentiel, ainsi que son contenu via des processus de dérivation et n'aura pas gardé de liens explicites avec les données de référence, sera confronté à de grandes difficultés lorsqu'il voudra intégrer les mises à jour fournies par le producteur. Il devra, en effet, interpréter les mises à jour délivrées afin d'en mesurer les conséquences sur son système, avant de pouvoir les intégrer. Cette interprétation peut être rendue difficile du fait de la diversité et de la complexité des traitements appliqués par l'utilisateur sur le référentiel.

Ces différents points viennent donc renforcer le besoin, déjà exprimé (cf. section 2.3.2.1), de disposer de modes de livraison spécifiques et dédiés à l'échange de l'information d'évolution entre les bases de données géographiques. Ces modes devront de surcroît permettre la livraison de mises à jour adaptées aux systèmes récepteurs, tant

dans le contenu que dans la forme. Au même titre qu'ils sont actuellement réalisés par [CEN, 1996 ; ISO, 1999 ; OGIS, 1999b] pour l'échange de données statiques, des efforts de normalisation devront être entrepris afin de standardiser ces échanges d'information dynamique.

Le panorama des problèmes entravant la mise à jour des systèmes d'information géographique que nous venons de dresser, a déjà permis de mettre en exergue un certain nombre d'éléments qui pourraient aider à leur résolution. Dans cette démarche de définition d'un mécanisme d'intégration des évolutions et de propagation des effets de celle-ci dans les bases de données, il convient de s'intéresser aux apports que peuvent fournir les solutions proposées actuellement pour aider à l'opération de mise à jour de l'information géographique. Ceci fait l'objet de la section suivante.

2.4 La mise à jour dans les systèmes d'information géographique

Même si elles n'adressent pas l'ensemble des problèmes que pose la mise à jour des bases de données géographiques et répondent bien souvent à des objectifs différents de ceux que nous cherchons à atteindre, les apports fournis par les solutions existantes sur la définition de notre mécanisme sont considérables.

Ainsi l'ensemble des travaux concernant la recherche des changements dans des couples d'images aériennes ou satellitaires et l'intégration de ceux-ci dans des bases de données vecteur [Huet et al., 1996 ; Jamet, 1998 ; Tellez et Servigne, 1997], constitue une première source d'apport importante. En effet, s'attachant plutôt à définir des processus automatiques de production et de maintenance de données géographiques de référence, ces travaux ont conduit au développement de nombreux algorithmes d'appariement, de plus en plus robustes, qui permettent de retrouver les éléments se correspondant dans chacune des images et ainsi de détecter les évolutions intervenues. Parallèlement, les travaux concernant l'appariement de données géographiques vecteur engagés au laboratoire COGIT depuis de nombreuses années [Lemarié, 1996 ; Lemarié et Raynal ; 1996 ; Bucaille, 1996 ; Devogele, 1997 ; Bel Hadj Ali, 1997 ; Devogele et al., 1998], s'ils ne se sont pas intéressés directement au problème de détection de changements dans les bases de données géographiques, peuvent aider à définir un mécanisme permettant d'isoler les évolutions. Ce mécanisme, comme nous l'avons déjà mentionné (cf. section 2.3.2.1), peut s'avérer utile voire indispensable à toute opération de propagation des mises à jour, en l'absence de systèmes d'identification ou de gestion de l'historique dans les bases mises en jeu.

Les travaux relatifs à l'implantation de la notion de temps dans les systèmes d'information géographique, constituent également un apport important pour la définition de notre mécanisme. En effet, les systèmes auxquels ils ont conduit fournissent un cadre formel et théorique dans lequel la maintenance des systèmes peut être réalisée de façon cohérente et donc participer à la définition d'une propagation de ses effets sans perte d'information ou obtention d'états incohérents. Basés sur les notions d'état (un état correspond à une configuration stationnaire d'un objet dans le temps) et d'événement (désignant la période durant laquelle l'objet est modifié), ces systèmes ont revêtu différentes formes :

- **L'estampillage temporel de données (time stamp)** [Langran et Chrisman, 1988 ; Langran, 1992] : Dans un tel système, chaque état des objets est conservé dans la base. Seul un attribut du type « date de création » ou « date de suppression », porté par chacun des objets, permet d'en identifier la version valide à une date donnée. Ces systèmes entraînent donc un fort taux de réplication des données et posent de nombreux problèmes de cohérence temporelle, la valeur des estampilles portées par chacun des objets étant, en effet, difficilement généralisable à un état complet de la base. Afin d'éviter les problèmes de réplication, d'autres systèmes plus sophistiqués, comme l'estampillage d'attributs [Snodgrass, 1992], permettant de ne pas dupliquer les parties invariantes des objets, ont été proposés [Jensen et Snodgrass, 1996], mais semblent difficiles à mettre en œuvre.
- **Le versionnement d'objets** : S'appuyant sur la construction d'arbres de version d'objets, ces systèmes évitent les problèmes de redondance posés par l'estampillage. Chaque objet dispose d'un numéro de version et seuls les attributs modifiés sont stockés dans ses différentes versions. Un mécanisme de gestion, interne à ce type de système, permet alors de déduire l'état complet d'une version d'un objet par « chaînage » des modifications apportées dans les versions précédentes. Au même titre que l'estampillage, ce système n'assure pas la préservation de la cohérence temporelle lors de l'extraction d'un état complet de la base.
- **Le versionnement de bases de données** : Ce système, qui étend le concept de version d'objets à l'ensemble d'une base de données et dont l'utilité dans les systèmes d'information géographique a été illustrée dans [Bauzer-Meideros et Jomier, 1994], permet non seulement de stocker, sans redondance, les différents états des objets mais aussi l'ensemble des relations qui les lient. Il permet ainsi d'assurer la cohérence entre les différentes versions des objets contenus dans les bases de données [Cellary et Jomier, 1990].

La gestion indépendante du temps et de la dimension spatiale des objets ne permettant pas de traduire au mieux, les changements que peuvent subir les entités géographiques du monde réel, de nombreux travaux ont visé à leur unification [Worboys, 1994 ; Cheylan et al., 1994 ; Claramunt et al., 1994 ; Hornsby et Egenhofer, 1997 ; Hornsby et Egenhofer, 1998]. Les différentes typologies des évolutions subies par les objets géographiques qui y ont été établies, constituent ainsi un apport important pour la qualification de l'information de mise à jour dans les bases de données géographiques et aident, de ce fait, à établir un processus d'intégration plus cohérent et plus facilement maîtrisable de ces changements dans les systèmes, car prenant en compte une information plus proche des évolutions réellement constatées dans le monde réel.

Les travaux concernant la mise à jour de vues dans les bases de données classiques [Scholl et al., 1991 ; Amer-Yahia et al., 1996], même s'ils adressent la maintenance de systèmes multi-représentations particuliers, constituent également une source d'apport intéressante pour la définition d'un mécanisme de mise à jour de la donnée géographique. En effet, les processus de propagation de changements sémantiques qu'ils mettent en œuvre entre des représentations différentes et parfois « éloignées » (en terme de niveaux d'abstraction) d'une même réalité, peuvent aider, pour partie, à la définition du mécanisme de d'intégration et de propagation des mises à jour que nous cherchons à établir.

Plus proche de nos préoccupations, le mécanisme de propagation des mises à jour défini dans [Kilpeläinen, 1994 ; Kilpeläinen, 1995 ; Kilpeläinen, 1997] fournit des indications intéressantes sur le type d'information à prendre en compte afin d'assurer une mise à jour cohérente. Ce mécanisme s'établit dans des systèmes multi-représentations particuliers. En effet, ces systèmes, appelés *incremental generalisation environment* (environnement de généralisation incrémental) sont constitués d'une base de données de référence (*geodata base*) et d'un ensemble de jeux de données dérivées de celle-ci. Les processus de dérivation qui ont conduit à l'établissement des différents jeux de données dérivés (*generalised output*) sont uniquement des processus de généralisation [Ruas, 1999] et sont supposés être formellement définis. Le mécanisme développé par Tiina Kilpeläinen n'adresse donc pas la mise à jour de systèmes constitués par ajout de couche d'information ou impliquant des processus semi-automatiques de dérivation. Néanmoins, la mise à jour y est opérée comme décrit ci-dessous, par généralisation locale liée des données de mise à jour (cette notion n'incluant pas seulement celle de proximité spatiale) :

« In an incremental generalisation environment the generalisation process is done completely for the whole geodata base only once. After the following update transactions to the geodata base, the old generalised output is also updated in an incremental way, which means that the generalisation process is performed only for the modules influenced by the updates » [Kilpeläinen, 1995].

La préservation de la cohérence dans de tels systèmes passe donc par une identification précise des *modules*, de la base de référence, concernés par l'opération de mise à jour. Bien qu'elle ne définisse pas clairement ce que recouvre le terme de *module*, on sent bien que celui-ci se doit d'inclure non seulement les objets modifiés mais aussi toute l'information contextuelle nécessaire à la génération d'une représentation à jour de ceux-ci dans les bases dérivées qui soit compatible avec le processus complet de généralisation. Cela implique donc, si l'on veut propager les mises à jour de façon cohérente, de pouvoir adapter les processus de dérivation mis en jeu. Une gestion de cette propagation basée uniquement sur l'utilisation de contraintes étendues d'intégrité spatiale ne permet, en effet, pas d'atteindre cet objectif, car elle rend difficile la propagation des modifications induites, comme le démontre [Kempainen, 1994].

Le mécanisme défini dans [van Wijgarden et al., 1997 ; Uitermark et al., 1998], qui permet la propagation des évolutions du bâti entre deux bases de données topographiques présentant des échelles différentes, constitue également une approche intéressante de ce que pourrait être un processus de mise à jour des systèmes multi-représentations. Il met, en effet, en exergue l'importance de l'utilisation de stratégies d'intégration des évolutions, afin de mener à bien l'opération de mise à jour. La stratégie de propagation qu'ils mettent en œuvre, s'établit comme suit :

1. Le contenu de la base de plus grande échelle est tout d'abord ramené à un état présentant l'actualité de la base de plus petite échelle.
2. Les correspondances entre les bâtiments présents dans les deux jeux de données sont ensuite établies, via l'utilisation d'un processus de superposition (*overlay* en anglais) [van Oosterom, 1994].
3. Les différences de spécifications entre les deux bases mises en jeu induisant l'établissement de liens incohérents, les correspondances sont alors vérifiées afin de les éliminer.

4. Les correspondances sont alors regroupées afin d'établir les équivalences entre groupes de bâtiments. Les mises à jour opérées sur la base de plus grande échelle sont filtrées afin de ne retenir que celles qui sont pertinentes vis à vis des spécifications de la base de plus petite échelle.
5. Les objets à jour, ainsi sélectionnés, sont alors généralisés.
6. Finalement, ces objets à jour et généralisés sont intégrés dans la base de plus petite échelle, à l'aide d'un algorithme de déformation élastique (rubber-sheeting ajustement algortihm)

Bien que limité à la mise à jour du bâti, ce mécanisme met en exergue l'importance de la définition et de l'utilisation de stratégies de propagation afin d'opérer une mise à jour la plus cohérente possible des bases de données géographiques. Les conflits d'intégration (comme, par exemple, les intersections éventuelles entre les bâtiments mis à jour et ceux n'ayant pas évolué) ne sont toutefois pas détectés. Aussi, si l'on veut réaliser une intégration totalement cohérente des mises à jour dans les systèmes multi-représentations, il faudra se doter de tels outils de détection et établir des systèmes permettant leur reprise.

Au vu des problèmes entravant la mise à jour des systèmes d'information géographique qui ont été identifiés dans la section 2.3 et des apports fournis par les solutions existantes, mis en exergue dans cette section, un mécanisme permettant de réaliser une intégration des mises à jour et une propagation des effets de celles-ci dans les bases de données géographiques utilisateur ou dérivées, qui soit la plus automatique possible et qui n'entraîne pas de dégradations de la cohérence et des acquis dans de tels systèmes, peut être défini. La section suivante présente sa structure et son fonctionnement d'ensemble. La description succincte des techniques que ce mécanisme met en œuvre, depuis l'extraction des évolutions jusqu'à la propagation des mises à jour dans les bases de données utilisateur ou dérivée, permettra de montrer que l'ensemble des problèmes que pose l'opération de mise à jour est adressé.

2.5 Le mécanisme de mise à jour des BDG

Définie en détail dans [Badard et Lemarié, 1999] (article fourni dans le Chapitre 4), la structure de ce mécanisme de mise à jour des bases de données géographiques est illustrée dans la Figure 2.3.

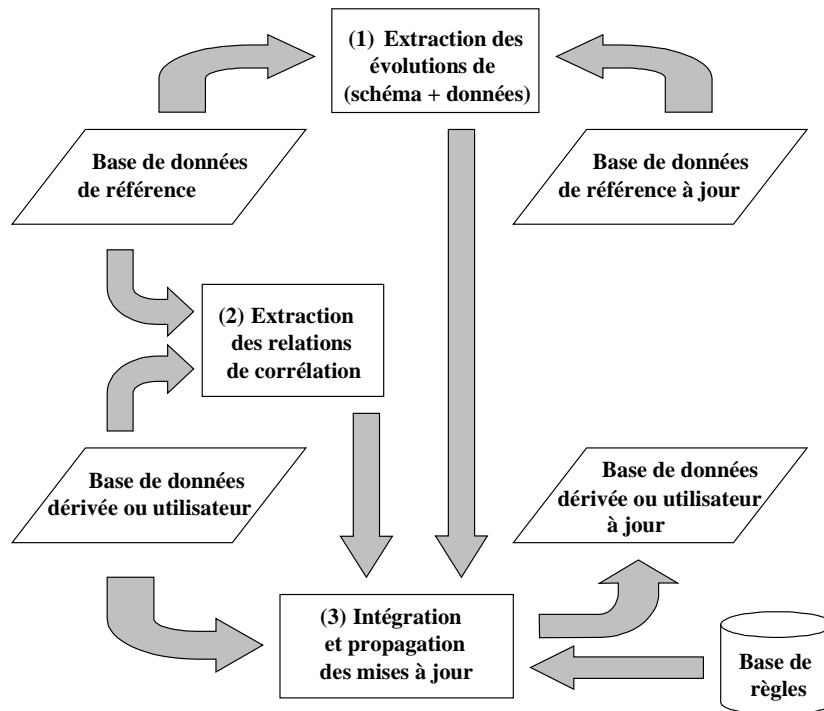


Figure 2.3 - Structure du mécanisme de mise à jour des bases de données géographiques

Ce mécanisme peut donc se décomposer en trois grandes étapes successives, s'appuyant chacune sur un mécanisme spécifique et s'étendant de l'extraction des évolutions dans le jeu de données de référence jusqu'à l'intégration et la propagation des effets de celles-ci dans les bases de données utilisateur ou dérivées. Ces étapes sont décrites dans les sections suivantes.

2.5.1 L'extraction des évolutions

Afin de répondre aux problèmes d'isolation et d'accès à l'information d'évolution (cf. 2.3.3.1 et 2.3.3.2) dans les bases de données de référence, un outil générique de détection et d'extraction automatique des mises à jour entre deux actualités d'une même base a dû être, tout d'abord, défini et développé. Il s'appuie sur des méthodes d'appariement de données géographiques qui permettent de retrouver dans deux représentations du monde réel, les entités ou groupes d'entités géographiques se correspondant. Il est à noter que ces deux jeux de données peuvent présenter des niveaux d'échelle, d'abstraction et de représentation très différents. Ces méthodes d'appariement de données géographiques sont basées sur un grand nombre d'algorithmes géométriques, topologiques et sémantiques, développés au laboratoire COGIT de l'IGN depuis plusieurs années [Lemarié, 1996 ; Lemarié et Raynal ; 1996 ; Bucaille, 1996 ; Devogele, 1997 ; Bel Hadj Ali, 1997 ; Devogele et al., 1998], ainsi qu'au cours de ce travail de thèse, permettant de traiter tous les types d'entités géographiques (à savoir : ponctuel, linéaire, surfacique).

Du fait de l'utilisation de telles méthodes, le mécanisme de détection et d'extraction des mises à jour développé permet de s'absoudre des problèmes liés à l'absence de système d'identification ou de gestion d'historique au sein des bases de données mises en jeu (cf. 2.3.1.1 et 2.3.3.1), systèmes qui peuvent, comme nous l'avons vu (cf. 2.3.3.1), être relativement difficiles à gérer et à maintenir.

L'établissement de ce mécanisme de détection et d'extraction de mises à jour a nécessité de surcroît la définition d'une typologie des évolutions spatio-temporelles que peuvent subir les entités géographiques modélisées au sein des bases de données, de façon à lui permettre de délivrer une information de mise à jour à la fois détaillée et proche des modifications opérées et opérables dans les systèmes d'information géographique concernant non seulement les données mais aussi les schémas des applications. Ce mécanisme rend ainsi possible la définition d'un processus d'intégration des évolutions dont les effets dans les bases de données utilisateur et dérivées sont plus appréhensibles et maîtrisables. Il apporte donc des éléments de réponse aux problèmes d'expression de l'information d'évolution (cf. 2.3.2.3) dans les bases de données géographiques et de changement de spécifications que peut opérer un producteur (cf. 2.3.2.1).

Ce mécanisme a, enfin, permis d'aider à la définition de nouveaux modes de livraison de l'information d'évolution pour la donnée géographique (notions de lots différentiels et d'évolution), répondant ainsi aux manques actuels de formats d'échange adaptés à ce type d'information (cf. section 2.4) et permettant également d'entrevoir la mise à disposition de la mise à jour sur les réseaux, comme Internet.

La définition détaillée, ainsi que le fonctionnement, de ce mécanisme seront précisés dans le chapitre 3 de ce mémoire. Les modes de livraison de l'information d'évolution qui ont été également définis, seront également présentés et discutés au sein de ce même chapitre.

2.5.2 L'extraction des relations de corrélation

Les problèmes de manque de structuration des bases de données utilisateur ou dérivées, exprimés dans la section 2.3.1.2, ont mis en exergue qu'un certain nombre de relations implicites entre données de référence et données ajoutées (ou dérivées), devaient être établies avant toute intégration des mises à jour, afin de permettre une propagation, sans perte d'information ni violation des règles d'intégrité (et notamment spatiales), dans les bases. Basé, en partie, sur les travaux de [Egenhofer et Herring, 1990 ; Egenhofer et al., 1994 ; Egenhofer et Mark, 1995], définissant un cadre formel des relations spatiales pouvant exister entre objets géographiques, et de [Lemon et Masters, 1997 ; Lemon, 1998], sur la gestion de celles-ci au sein d'un système cadastral, cet ensemble minimal de relations, que nous avons nommées *relations de corrélation* et qui rendent possible le processus de mise à jour des bases de données géographiques, a été établi dans [Badard, 1998a ; Badard, 1998b]. Cet ensemble peut se décomposer en 4 principaux types :

- **Les relations de composition**

Ce sont les relations de composition telles que nous l'entendons dans le sens commun, par exemple un département est composé de communes. Elles sont souvent déjà présentes dans les bases de données géographiques, mais l'ajout de couches d'information dans les systèmes peut entraîner l'apparition de nouvelles relations de ce type. Elles doivent donc être retrouvées, afin de permettre la préservation d'une partie des contraintes d'intégrité spatiale au cours du processus de propagation de mise à jour.

- **Les relations de correspondance**

Ce sont les relations qui lient les objets ou groupes d'objets représentant les mêmes phénomènes du monde réel. Ces relations permettent donc de retrouver les objets de la base utilisateur ou dérivée qui s'apparentent aux données contenues dans le référentiel, et qui subiront ainsi directement les mises à jour de ce dernier.

- **Les relations de nature géométrique**

Ces relations traduisent l'ensemble des relations spatiales qui peuvent exister entre les objets géographiques stockés dans les bases, à savoir les relations de partage de géométrie, de connexion, d'intersection et de proximité. Elles permettent de retrouver les objets qui subissent des modifications induites par des mises à jour du référentiel intégrées dans la base de données utilisateur ou dérivée. Ces relations rendent possible la préservation des contraintes d'intégrité spatiale au cours du processus de propagation des effets de la mise à jour.

- **Les relations de dépendance**

Ce sont des relations « classiques » de dépendance employées dans le cas d'informations déduites par calcul (par exemple, valeur d'un attribut calculée à partir de celles d'autres objets). Néanmoins, dans le cas des bases de données géographiques, la géométrie peut être considérée comme un attribut. Aussi, tout changement de la géométrie des objets peut nécessiter la modification d'information sémantique portée par les objets (exemple : valeur de la densité d'habitation d'une zone). Ces relations permettent donc de préserver les contraintes d'intégrité d'ordre sémantique.

Ces relations traduisent en quelque sorte les liens entre les objets géographiques qu'a ou qu'aurait induit un processus de dérivation, si tant est qu'il y en ait eu un. Elles sont établies non seulement entre les bases de données (dans leurs versions anciennes) mais également à l'intérieur du jeu de données utilisateur ou dérivé.

L'extraction de ces relations de corrélation a nécessité la définition et le développement d'un mécanisme dédié, dont la structure a été établie au cours d'un stage réalisé au sein du laboratoire COGIT et peut être trouvée dans [Bonnani, 1998]. Ce mécanisme s'appuie partiellement sur le processus d'appariement de données géographiques mentionné dans la section précédente. Seule l'utilisation de seuils de recherche et de détection adaptés, ainsi que de contraintes sémantiques différentes portant sur les classes d'objet à mettre en relation viennent le détourner de son but premier. De nombreux autres algorithmes permettant, notamment, l'extraction des relations de nature géométrique et de correspondance (méthodes d'appariement multi-échelles permettant de corréler les données géographiques autres que les réseaux, ces derniers ayant été déjà largement traités dans [Devogele, 1997]) ont été également développés au cours de ce travail de thèse. La Figure 2.4 fournit un exemple d'extraction des relations de partage de géométrie retrouvées entre les thèmes administratif et hydrographique de la BDCarto® [IGN, 1999].

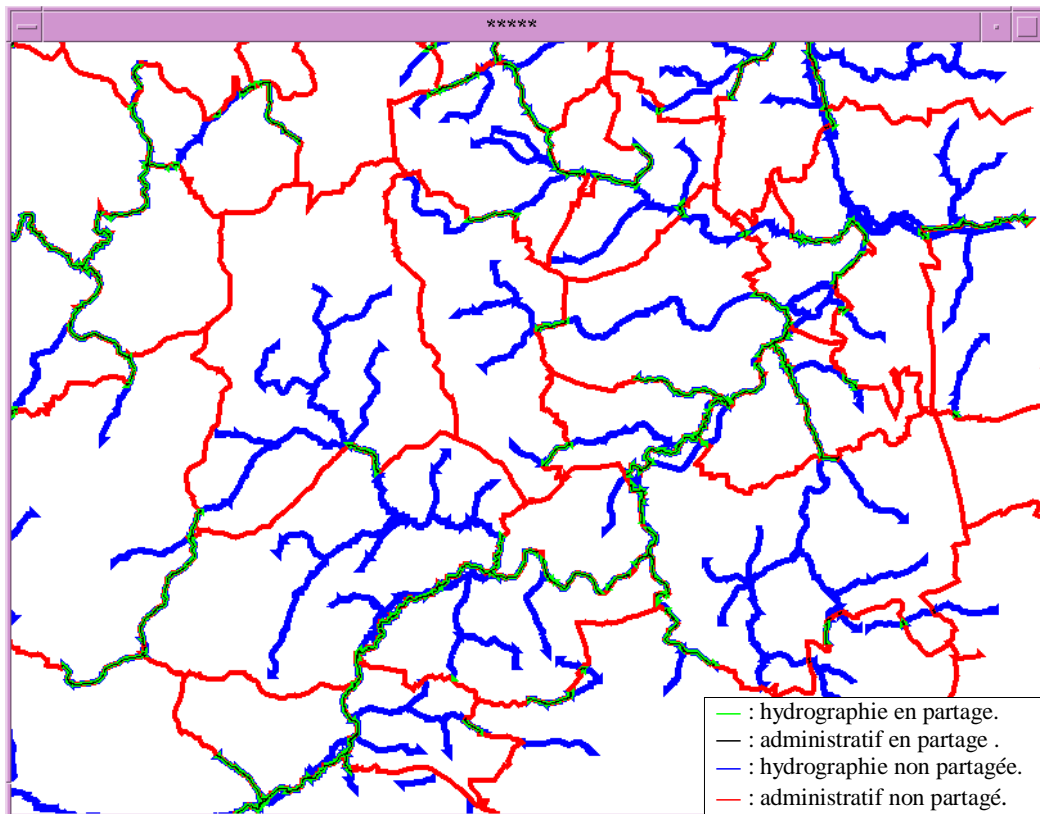


Figure 2.4 - Exemple de relations de partage de géométrie retrouvées entre les thèmes administratif et hydrographie de la BDCarto®

Du fait de l'utilisation au sein de ce mécanisme de méthodes d'appariement et de recherche des relations de corrélation basées sur des seuils de tolérance, les problèmes d'ajout d'information de niveaux de qualité divers (cf. 2.3.1.3) sont donc pris en compte. Bien loin de rétablir la cohérence complète et qui n'existe parfois que partiellement dans les bases de données utilisateur ou dérivées, ce mécanisme permet au moins de rendre possible l'intégration des mises à jour (via l'extraction des objets homologues) et d'en propager les effets (via la détection des objets en relation avec les données du référentiel) tout en conservant les contraintes d'intégrité existantes. Il peut permettre de surcroît, via la mise en œuvre d'algorithmes dédiés de correction de la topologie, de rétablir la cohérence dans de tels systèmes non seulement à l'intérieur d'une même couche géométrique, tel que présenté dans [Ubeda et Egenhofer, 1997], mais aussi entre les différents thèmes.

Dans le but d'automatiser ce processus d'extraction et de capitaliser la connaissance acquise sur les jeux de données mis en (cor)relation, des métadonnées spécifiques décrivant la nature des entités, les échelles, et les valeurs des seuils employés lors de la détection ont été également définies [Bonnani, 1998].

Ce processus s'appuyant très largement sur celui de l'appariement de données géographiques, le lecteur est invité à se référer au chapitre 3 de ce mémoire, afin de connaître les principes de son fonctionnement.

2.5.3 L'intégration et la propagation des effets de la mise à jour

Ce dernier processus constitue le cœur du mécanisme d'intégration et de propagation de la mise à jour dans les systèmes utilisateur ou dérivés. Comme illustré dans la Figure 2.3, il s'appuie sur les relations de corrélation précédemment extraites et qui vont permettre de propager les effets des mises à jour détectées dans le jeu de données de référence.

Afin de contrôler la propagation et permettre une mise à jour complète et cohérente des bases de données géographiques, une base de règles est mise en œuvre. Cette base de règles tient compte des spécifications et des métadonnées disponibles (schéma, échelles, résolution, etc.) et conditionne le déclenchement des algorithmes appropriés, permettant la mise à jour des entités géographiques stockées dans les bases de données. Elle permet ainsi d'établir la stratégie de mise à jour qui doit être appliquée et de définir l'ordre dans lequel les évolutions doivent être intégrées et propagées, tel que décrit dans [Cordier et Siegel, 1994] pour les bases de données classiques, afin de ne pas entraîner de perte d'information ou l'obtention d'états incohérents.

Dans le but de permettre la gestion de conflits potentiels, non prévus, mais néanmoins détectés par le mécanisme d'intégration et de propagation au cours du processus de mise à jour, la base de règles a été conçue de façon à offrir le maximum d'ouverture. Elle rend ainsi possible et facile, l'ajout de règles supplémentaires et spécifiques pour la gestion de ces conflits. Les conflits non résolus in fine, malgré ces révisions de la base de règles, sont identifiés et listés par le mécanisme et peuvent être alors résolus de façon interactive. Des outils dédiés ont été développés pour mener à bien cette tâche.

La définition détaillée ainsi que le fonctionnement de ce mécanisme d'intégration et de propagation des effets des mises à jour seront précisés dans le chapitre 4 de ce mémoire. La mise en œuvre de la base de règles ainsi que du modèle de gestion de la propagation sur lesquels il s'appuie, y seront également discutés et illustrés.

2.6 Cas d'étude

Afin de valider le mécanisme que nous venons de présenter dans la section précédente, deux tests concernant des bases de données géographiques produites par l'Institut Géographique National et mettant en jeu des processus de dérivation différents ont été entrepris et menés à bien. Ils sont présentés dans la section suivante.

Ces cas d'étude ont concerné la propagation des mises à jour de la BDCarto sur Route500 et de la BDTopo sur la BDTopo25. Le premier test adresse les problèmes de mise à jour d'une base de données dérivée issue d'un processus de généralisation, induisant des écarts d'échelles importants entre les jeux de données. Le deuxième s'intéresse aux difficultés liées à un processus de rédaction cartographique, impliquant un changement de représentation conséquent entre les bases.

Ces tests ont été réalisés sur GéO₂, le prototype de système de gestion de bases de données géographiques développé au laboratoire COGIT et au sein duquel l'ensemble des développements liés aux travaux de recherche engagés depuis plusieurs années, a été intégré.

Les deux cas d'étude investigués, ainsi que la plate-forme de développement sur laquelle ils ont été menés sont succinctement décrits dans les sections suivantes.

2.6.1 Propagation des mises à jour de la BDCarto vers Route500

La BDCarto® [IGN, 1999] est une base de données géographiques couvrant l'ensemble du territoire pour des échelles comprises entre le 1 : 100 000 et le 1 : 500 000. Elle est adaptée à la cartographie de synthèse et aux applications d'études de projet, de gestion d'infrastructures et d'aménagement du territoire. Elle a un rôle de référentiel au niveau départemental et régional, facilitant le partage d'informations entre les différents acteurs d'une même collectivité territoriale, et sur lequel chaque utilisateur peut rattacher les données de son domaine thématique.

Elle est constituée à partir des cartes au 1 : 50 000 de l'IGN (scannage des cartes, vectorisation puis structuration topologique) et d'imagerie spatiale SPOT. Sa précision est de l'ordre de la vingtaine de mètres.

Les éléments du terrain sont regroupés en 9 thèmes (réseau routier, réseau ferré, hydrographie, franchissement, unité administrative, habillage, toponymes, équipement, occupation du sol). A chaque thème correspond une couche géométrique de type graphe planaire. La mise à jour est réalisée thème par thème à des rythmes très différents (tous les 6 mois pour le routier, tous les 10 ans pour l'occupation du sol). La cohérence n'est assurée qu'entre les thèmes routier, franchissement et ferré (partiellement avec l'hydrographie).

Route500® [IGN, 1998] est une base données à vocation essentiellement routière, contenant les 500 000 kilomètres de réseau national classé. Elle est obtenue à partir de la BDCarto, via un processus de dérivation impliquant une part importante de reprises manuelles. Les éléments du réseau routier de la BDCarto sont tout d'abord sélectionnés suivant des critères portant principalement sur les valeurs d'attributs. Les éléments sélectionnés, connectés entre eux, disposant d'un contenu sémantique identique, et compris entre deux carrefours sont ensuite fusionnés, lissés puis filtrés (via un processus du type [Douglas et Peucker, 1973]). Les valeurs des paramètres d'entrée impliqués dans les algorithmes de lissage et de filtrage dépendant essentiellement de la longueur et du type de la route (autoroute, nationale, etc.). L'ensemble du résultat obtenu est alors repris entièrement manuellement par un opérateur pour correction et complément (le processus de sélection induisant de nombreuses erreurs), afin notamment d'assurer la continuité du réseau routier, la logique de desserte et l'exhaustivité de la base.

La Figure 2.5 illustre la différence de contenu des thèmes routiers de la BDCarto et de Route500.

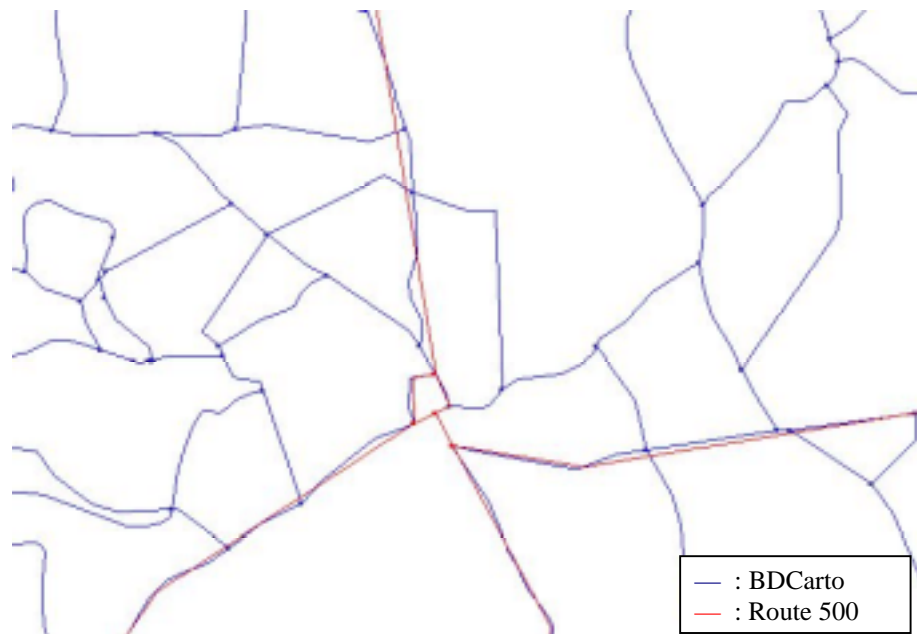


Figure 2.5 - Différences de contenu des thèmes routiers BDCarto et Route500

Le test de propagation des mises à jour de la BDCarto vers Route500 a porté sur l'étendue du département des Côtes d'Armor (22) avec une différence d'actualité entre les bases de 4 ans (version 1 de 1994 et version 2 de 1998), soit en terme de nombre d'objets géographiques :

	1994	1998
<i>BDCarto</i>	64303	63709
<i>Route500</i>	14436	5774

Tableau 2.1 - Différences de contenu en nombre d'objets entre la BDCarto et Route500

La valeur mentionnée pour la version de Route500 datant de 1998, correspond au nombre d'objets effectivement contenus dans la version produite par l'IGN de ce jeu de données. En effet, afin d'évaluer les résultats obtenus, nous avons dû disposer de cette version de Route500. L'écart de valeur entre les deux actualités de Route 500 est dû au fait que ces deux jeux de données ne présentaient pas la même emprise d'extraction.

Les résultats concernant l'extraction des mises à jour entre les deux actualités de la BDCarto seront fournis dans le chapitre 3 (cf. section 3.4.1) et ceux, relatifs à l'intégration et la propagation de ces évolutions dans Route500 seront donnés dans le chapitre 4 (cf. section 4.4.1).

2.6.2 Propagation des mises à jour de la BDTopo vers la BDTopo25

La BDTopo® [IGN, 1997a] fournit un système de référence de localisation pour les applications s'étendant du territoire d'une commune à celui d'un département. Elle se caractérise principalement par sa description détaillée de l'environnement, par sa précision de l'ordre d'un mètre et sa structure topologique de réseau. Son contenu

correspond globalement au contenu de la carte au 1 : 25 000. Elle permet des sorties graphiques du 1 : 5 000 au 1 : 25 000. La source géométrique principale est une saisie photogrammétrique (photographies aériennes). Elle est produite par stéréo-restitution. Des compléments issus d'un relevé de terrain sont également réalisés de façon systématique, ils concernent principalement les zones masquées sur les photographies aériennes et les zones où une validation sur le terrain est nécessaire.

La BDTopo est séparée en deux couches géométriques indépendantes :

- l'altimétrie (ou alti) qui regroupe les courbes de niveau et points cotés,
- la planimétrie (ou plani) qui regroupe tous les autres objets.

La BDTopo a été modélisée en privilégiant le concept de classe sur celui d'attribut. Elle est donc composée d'un grand nombre de classes qui ont peu d'attributs et de relations. Elle est mise à jour tous les 3 à 9 ans.

La BDTopo25® [IGN, 1997b] est la base de donnée géographiques qui sert à l'édition des cartes papier au 1 :25 000, issues de la filière numérique (série TOP25 de l'IGN). Elle est dérivée de la BDTopo, via un processus presque entièrement automatique. Ce processus consiste d'abord en une série de traitements de sélection basés sur la sémantique et la géométrie des objets afin de remplir le schéma spécifique de la BDTopo25, d'algorithmes de généralisation cartographique et de symbolisation visant à rendre les géométries conformes aux spécifications de la carte au 1 : 25 000. Une phase de reprise interactive est au final réalisée afin d'améliorer la lisibilité de la carte. Elle consiste principalement à déplacer ou à éliminer les objets dont les symbolisations se recouvrent.

La Figure 2.6 illustre la différence de représentation entre la BDTopo et la carte au 1 :25 000, issue de la filière numérique (série TOP25 de l'IGN).



Figure 2.6 - Différences de représentation entre la BDTopo (à gauche) et la carte au 1 : 25 000 (TOP25, ©IGN, à droite)

Le test de propagation des mises à jour de la BDTopo vers la BdTopo25 a porté sur une partie de la feuille d'Angers (Département du Maine-et-Loire, 49) avec une différence d'actualité entre les bases de 2 ans (version 1 de 1994 et version 2 de 1996), soit en terme de nombre d'objets géographiques :

	1994	1996
<i>BDTopo</i>	9079	9891
<i>BDTopo25</i>	8299	8940

Tableau 2.2 - Différences de contenu en nombre d'objets entre la BDTopo et la BDTopo25

La valeur mentionnée pour la BDTopo25 datant de 1996, correspond au nombre d'objets effectivement contenus dans la version produite par l'IGN de ce jeu de données. En effet, afin d'évaluer les résultats obtenus par le mécanisme de mise à jour, nous avons dû disposer de cette version de la BDTopo25.

Les résultats relatifs à l'extraction des mises à jour entre les deux actualités de la BDTopo seront fournis dans le chapitre 3 (cf. section 3.4.2) et ceux, concernant l'intégration et la propagation de ces évolutions dans la BDTopo25 seront donnés dans le chapitre 4 (cf. section 4.4.2).

2.6.3 Présentation de GéO₂

Le prototype de SGBD géographiques GéO₂ [David et al., 1993c ; Raynal et al., 1995] mis au point au laboratoire COGIT de l'IGN et intégrant l'ensemble des développements liés aux travaux de recherche engagés dans le passé en terme de gestion de base de données, a été choisi comme plate-forme pour mener à bien les tests. Il repose sur le SGBD orienté objet du commerce O₂ [O₂, 1991 ; Bancilhon et al., 1992].

L'orienté objet est avantageux pour les BDG [David et al., 1993a ; David et al., 1993b ; Abrantes, 1996 ; Worboys et al., 1990 ; Frank et Egenhofer, 1988], car il autorise une meilleure modélisation du monde réel par des objets complexes et la définition de types correspondant aux primitives géométriques. Qui plus est, il permet d'éviter les nombreuses jointures nécessaires pour manipuler les données avec un modèle relationnel.

GéO₂ intègre la plupart des fonctionnalités des SIG. Le modèle de données de GéO₂ [David et al., 1993c] a été défini au laboratoire COGIT. Il a été construit afin que les objets géométriques (points, lignes, surfaces) soient dissociés des objets géographiques qu'ils représentent (les tronçons de routes, les habitations,...). Chaque classe d'objets géographiques a un attribut de localisation du type Geometry. Ce type est défini comme le plus petit Type Abstrait de Données (TAD) [Bouillé, 1987] incluant les points, les lignes et les surfaces [David, 1991]. Les objets d'une même classe géographique peuvent donc avoir des géométries de dimensions différentes. Réciproquement, les objets géométriques connaissent les objets géographiques qu'ils représentent.

Les géométries sous GéO₂ sont regroupées en couches géométriques. Trois types de couches géométriques sont définies : les couches « Spaghetti », les couches « réseau planaire » et les couches « carte topologique » [Scholl et al., 1996 ; Laurini et Thompson, 1992 ; Denégre et Salgé, 1996]. Ainsi, plusieurs niveaux de représentation de la topologie sont possibles pour les géométries. Une base GéO₂ peut contenir plusieurs domaines, chaque domaine étant une représentation. Les géométries d'un domaine sont groupées dans une ou plusieurs couches. De plus, GéO₂ autorise le stockage de métadonnées dans un dictionnaire. Par contre, GéO₂ ne gère pas la troisième dimension. Le chargement des données dans GéO₂ est réalisé domaine par

domaine à partir de lots au format FEIV (Format d'Echange Interne Vecteur) [Richard, 1993], qui est le format défini au départ par Pascal Etienne en 1990 et utilisé en interne à l'IGN, pour l'archivage des bases de données géographiques vecteurs. Au cours de ce travail de thèse, afin de pouvoir livrer les résultats d'études connexes dans lesquelles nous avons été impliqués, des développements ont dû être entrepris afin de rendre Géo₂ capable d'exporter des données dans des formats classiques de logiciels de SIG du marché. Ainsi, il est maintenant possible d'exporter des données au format Shapefile (ESRI²), MIF/MID (Map Info³) et Export Géoconcept (Géoconcept S.A.⁴)

² Voir : <http://www.esri.com>.

³ Voir : <http://www.mapinfo.com>.

⁴ Voir : <http://www.geoconcept.com>.

Chapitre 3

Extraction des mises à jour et diffusion de l'information d'évolution

3.1 Introduction

Ce chapitre présente de façon détaillée le processus d'extraction des évolutions entre deux versions d'une même base de données géographiques (cf. 3.2.1) que nous avons mis au point et sur lequel s'appuie le mécanisme de mise à jour proposé. Ce processus constitue un préalable indispensable à toute propagation des mises à jour dans les systèmes. L'information qu'il délivre a permis la définition de modes de livraison (notions de lots différentiels et de lots d'évolution), permettant de mettre à disposition l'information d'évolution pour les bases de données géographiques. La description détaillée de ceux-ci et de leurs avantages et inconvénients, sera fournie également dans ce chapitre (cf. 3.2.1 et 3.2.2). Un certain nombre de compléments et de questions connexes que met en exergue le processus d'extraction sera ensuite fourni au lecteur (cf. 3.3), afin d'en compléter la compréhension des forces et faiblesses. Enfin, les résultats des extractions de mise à jour, opérées sur les différents cas d'étude entrepris, seront donnés afin d'illustrer son fonctionnement (cf. 3.4).

3.2 Contributions

Les articles référencés ci-dessous constituent la contribution apportée à la résolution du problème d'isolation, d'accès et de mise à disposition de l'information d'évolution dans les bases de données géographiques. Ces articles sont fournis dans les pages qui suivent.

3.2.1 Mécanisme d'extraction des mises à jour et première proposition pour leur échange (lots différentiels)

[Badard, 1998c] Extraction des mises à jour dans les BDG – De l'utilisation de méthodes d'appariement. Revue internationale de géomatique, vol. 8, n° 1-2, Hermès, Paris, 1998, pp. 121–147.

3.2.2 Echange de l'information d'évolution via l'utilisation de XML (lots d'évolution)

[Badard et Richard, 2000a] T. Badard et D. Richard. Using XML for the exchange of updating information between geographical information systems. To appear in Computers, Environment and Urban Systems (CEUS), Elsevier Science, Amsterdam, 2000, 12 pages.

Extraction des mises à jour dans les BDG

De l'utilisation de méthodes d'appariement

On the retrieval of updates in geographic databases by geographic data matching tools

Thierry Badard

*Institut Géographique National
Service de la Recherche/Laboratoire COGIT
2 à 4, avenue Pasteur
94165 – Saint Mandé Cedex, France
Email : Thierry.Badard@ign.fr*

RÉSUMÉ. *L'intégration dans une base de données géographiques (BDG), des mises à jour délivrées par un producteur de données de référence est encore problématique. Une des causes principales en est l'impossibilité d'isoler facilement cette information dans les nouveaux jeux de données fournis par le producteur. Cet article s'attache donc à proposer, dans un premier temps, un mécanisme générique de détection et d'extraction automatique des mises à jour afin de répondre à ce problème. Celui-ci s'appuie sur des méthodes d'appariement de données géographiques et est donc indépendant de toutes notions d'identifiant et de gestion d'historique dans les bases mises en jeu. La possibilité d'utiliser ce mécanisme dans la définition de nouveaux modes de livraison de l'information de mise à jour sera ensuite étudiée et la notion de calque différentiel sera notamment définie.*

ABSTRACT. *The integration in geographic databases of the updates delivered by a producer of reference data sets is still problematic. One of the main reasons is that the detection of updates in the new version of the database may be impossible to process. The purpose of this paper is thus to provide a generic tool for the automatic retrieval of the updates in geographic databases in order to make their integration easier. This mechanism is based on geographic data matching tools. Thus the implementation of systems of identifiers or time management is not necessary in the databases. The use of this process in the definition of new delivery modes of updates for geographic databases is also discussed.*

MOTS-CLÉS : *Mise à jour, appariement, évolutions spatio-temporelles, différentiels*
KEY WORDS : *Update, geographic data matching, spatio-temporal evolutions, versioning*

1. Introduction

De nos jours, un nombre croissant d'organisations font appel à des producteurs de données géographiques de référence afin d'asseoir, à moindre coût, leur propre système d'information géographique (SIG). Cependant, pour pouvoir répondre au mieux aux missions dont ils ont la charge, ces utilisateurs de bases de données de référence ont un besoin évident de mises à jour de la part du producteur. Mais, à l'heure actuelle, l'intégration de ces évolutions dans les bases de données utilisateur pose de nombreux problèmes. Une taxonomie complète de ceux-ci a d'ailleurs été établie dans [BAD 98].

Cependant, une des principales causes de ces problèmes est le faible nombre de modes de livraison mis en œuvre pour délivrer l'information de mise à jour. En effet, les évolutions d'un jeu de données de référence sont le plus souvent fournies par le producteur sous la forme de la base de données entière mise à jour. Le taux d'évolution d'une base de données géographiques étant communément estimé à environ 10% par an [RAY 96], les mises à jour se retrouvent ainsi disséminées dans un volume important de données n'ayant pas évolué. L'utilisateur désireux d'intégrer ces modifications dans son système d'information a alors le choix entre deux alternatives. Il peut soit remplacer entièrement l'ancienne version de son jeu de données de référence par la nouvelle, soit extraire par lui-même les évolutions subies par les entités géographiques modélisées dans la base de données, par comparaison des deux états de celle-ci. Cependant, la première solution peut entraîner des pertes importantes d'information ou l'obtention d'états incohérents de la base utilisateur après intégration des mises à jour. La deuxième solution, quant à elle s'avère difficile à mettre en œuvre si un système fiable et stable d'identifiants ou de gestion de l'historique n'est pas implanté dans les bases de données mises en jeu.

Cet article s'attache donc à fournir un outil générique de détection et d'extraction automatique des mises à jour entre deux actualités d'une même base données, préalable indispensable à une intégration plus facile et plus maîtrisable de ces évolutions dans une base utilisateur. Ce mécanisme s'appuie sur des méthodes d'appariement de données géographiques qui seront détaillées dans la section 2. Une typologie des évolutions spatio-temporelles que peuvent subir les entités géographiques modélisées dans les bases sera ensuite définie. Son utilisation dans l'analyse des liens d'appariement établis afin d'extraire l'information de mise à jour sera alors également détaillée dans la section 3. Finalement, la section 4 s'attachera à étudier en quoi ce mécanisme peut aider à la définition de nouveaux modes de livraison de l'information de mise à jour. La notion de calque différentiel sera notamment définie et illustrée par les résultats d'un test mené sur des données du réseau routier de la BDCarto®.

2. L'appariement de données géographiques

L'appariement de données géographiques est à rapprocher de la notion de « conflation » en anglais définie dans [SAA 88], [LAU 92]. Il consiste à établir des liens de correspondance entre des ensembles d'entités géographiques symbolisant les mêmes phénomènes du monde réel dans deux représentations de celui-ci [LUP 87], [LEM 96b], [LEM 98]. Il est à noter que ces deux représentations peuvent présenter des échelles et/ou niveaux d'abstraction et/ou de représentation très différents. La figure 1 ci-dessous illustre d'ailleurs le résultat de ce processus entre deux jeux de données ayant des échelles différentes.

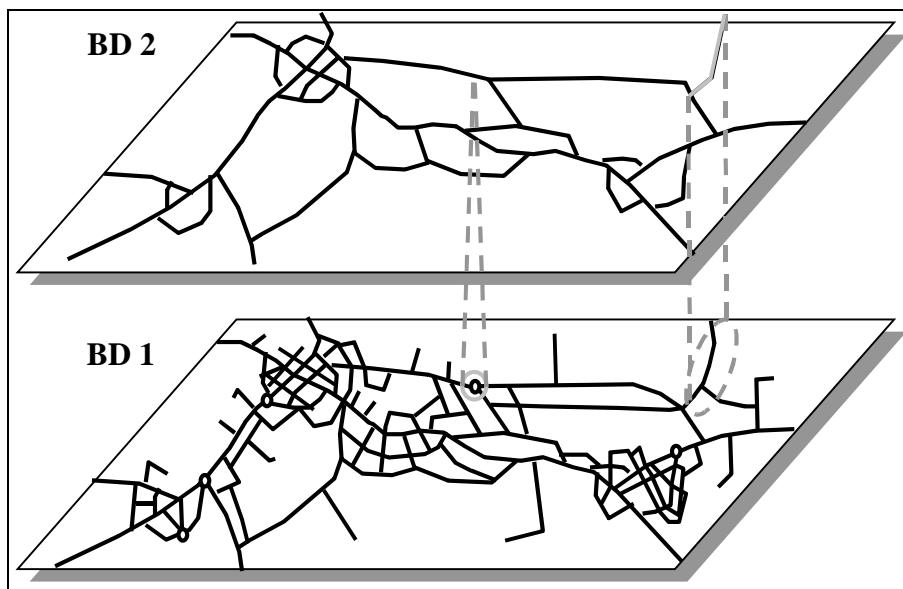


Figure 1. Illustration du mécanisme d'appariement

Ainsi, le rond point mis en valeur dans la base de données BD1 correspondra au nœud routier présent de la BD2. Le tronçon routier de la BD1 sera mis en correspondance avec l'élément de réseau routier de la BD2, ceci malgré le fait que des tronçons connectés soient absents dans cette dernière du fait de la différence d'échelles séparant ces deux jeux de données.

Un tel processus a été développé depuis plusieurs années au laboratoire COGIT et s'appuie sur un grand nombre d'algorithmes exploitant tous les niveaux de l'information géographique (à savoir : sémantique, géométrique et topologique). Il peut actuellement se décomposer selon le schéma suivant (cf. figure 2) :

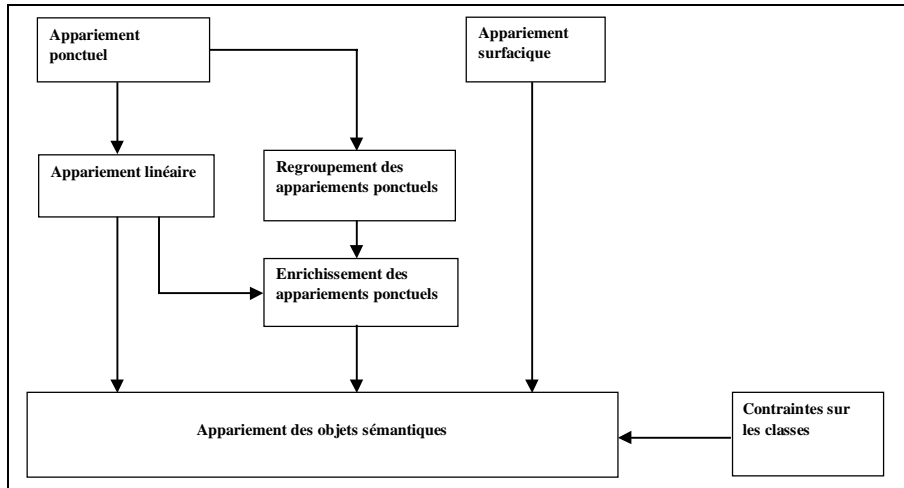


Figure 2. Fonctionnement général du module d'appariement

Ce processus adopte donc un comportement ascendant, comme l'illustre la figure 3. Il réalise tout d'abord l'appariement des primitives géométriques des objets avant de remonter aux objets (sémantiques) simples eux-mêmes et donc d'établir l'appariement entre ceux-ci. Les objets complexes, c'est à dire définis à partir des objets simples, ne disposent donc pas de géométrie propre. La géométrie de ces derniers est en fait la réunion de celles des objets simples les composant. Leur appariement résultera donc de celui des objets simples.

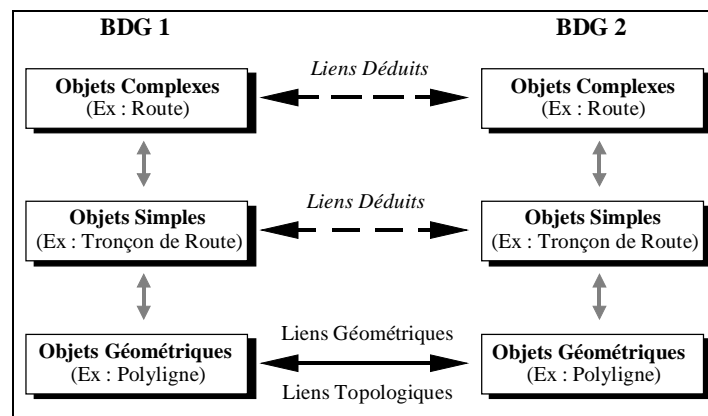


Figure 3. Comportement ascendant du processus d'appariement

Il est à noter que ce processus est quasiment entièrement automatique. Seule la phase d'enrichissement des appariements ponctuels nécessite une reprise interactive et guidée pour les appariements en conflit détectés par le processus. Le détail des

différentes étapes du processus d'appariement est donné dans les sections suivantes.

2.1. Appariement des primitives

Cet appariement est générique et ne dépend aucunement du schéma des bases de données à traiter, il utilise l'ensemble des primitives géométriques présentes dans les deux bases de données. Il se décompose en trois parties distinctes correspondant aux différents types de primitives géométriques que l'on peut rencontrer, à savoir : ponctuelles, linéaires et surfaciques.

Les méthodes utilisées sont géométriques et topologiques, la sémantique n'intervient pas à ce niveau car elle n'est de toute façon pas portée par les primitives géométriques. L'appariement des primitives surfaciques est totalement indépendant de l'appariement des primitives linéaires et ponctuelles. Par contre, il y a des interactions entre l'appariement ponctuel et l'appariement linéaire.

2.1.1. Appariement des primitives ponctuelles

L'utilisateur définit tout d'abord une précision pour la recherche des appariements ponctuels, c'est-à-dire une distance (en terme de distance euclidienne) de recherche des nœuds. Ce choix est en général fait en fonction des échelles des jeux de données à traiter et de la connaissance de l'erreur quadratique moyenne sur la position des éléments présents dans ces bases. Les primitives ponctuelles sont alors classées suivant le nombre décroissant d'arcs sortant et entrant. Ce classement est notamment utile pour l'appariement d'éléments de réseaux (ex : routiers, hydrographiques, lignes électriques, ...).

On se fixe ensuite une base de référence (l'autre base de données étant appelée alors base de comparaison) qui va permettre de connaître l'ordre dans lequel on va suivre l'appariement des nœuds. Pour chaque nœud de la base de référence, on recherche donc l'ensemble des nœuds de l'autre base situés dans la zone d'imprécision définie par le seuil (cf. figure 4, gauche). A partir de l'ensemble des nœuds trouvés, on vérifie ensuite un certain nombre de critères géométriques et topologiques afin de déterminer, dans cet ensemble de nœuds candidats à l'appariement, quels sont ceux réellement appariables. Ainsi, un test d'égalité géométrique (cf. figure 4, droite) est notamment réalisé et le nombre de primitives linéaires entrant et sortant de chaque nœud est vérifié.

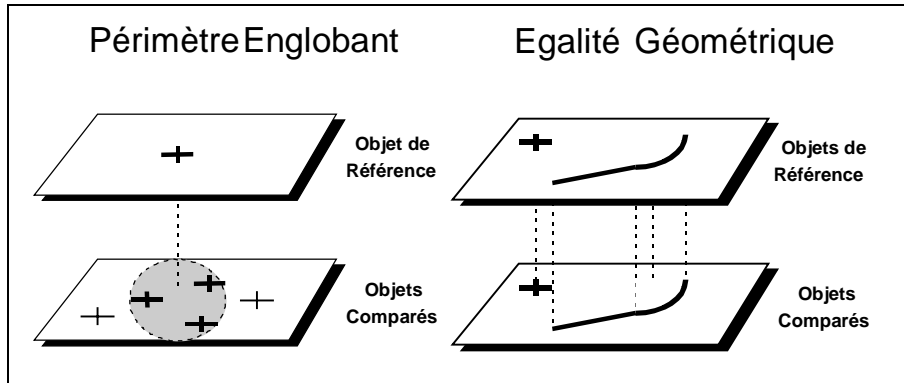


Figure 4. Méthodes employées lors l'appariement des primitives ponctuelles

L'appariement des primitives ponctuelles terminé, on passe à celui des objets géométriques linéaires.

2.1.2. Appariement des primitives linéaires

Cet appariement implique des algorithmes propres, inspirés de ceux décrits dans [ABB 94] et basés notamment sur la distance de Hausdorff. Il s'agit d'une distance permettant à la fois de mesurer des écarts de position mais aussi de forme entre les objets. Elle est définie comme suit (cf. figure 5) :

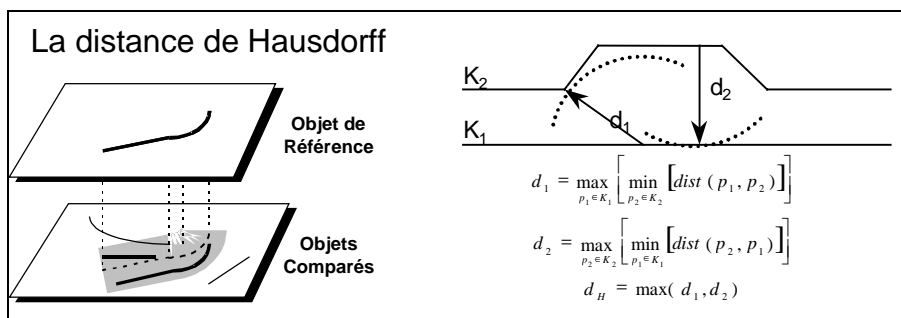


Figure 5. Définition de la distance de Hausdorff

Il est cependant à noter que le choix d'une base de référence ayant été réalisé au préalable, seule une composante de la distance de Hausdorff est utilisée. On recherche donc pour chaque primitive linéaire de la base de référence, les objets présents dans l'autre base à une distance (en terme de distance de Hausdorff cette fois ci) inférieure à un seuil fixé par l'utilisateur. Ce choix, comme précédemment,

repose sur la connaissance des échelles et de l'erreur quadratique moyenne sur la position des éléments des jeux de données mis en jeu.

Cependant, cet appariement utilise également les liens d'appariement générés lors de l'appariement ponctuel. Les appariements ponctuels sont alors parcourus et analysés un à un puis on en déduit dans les deux bases les arcs entrant et sortant de ces nœuds appariés, en utilisant notamment la notion d'angle résultant (cf. figure 6, gauche). Ceci implique donc le choix d'un angle seuil au-dessus duquel un arc sera rejeté. Enfin, pour déterminer des liens entre primitives linéaires manquants, un algorithme de parcours de graphe topologique est également mis en œuvre (cf. figure 6, droite).

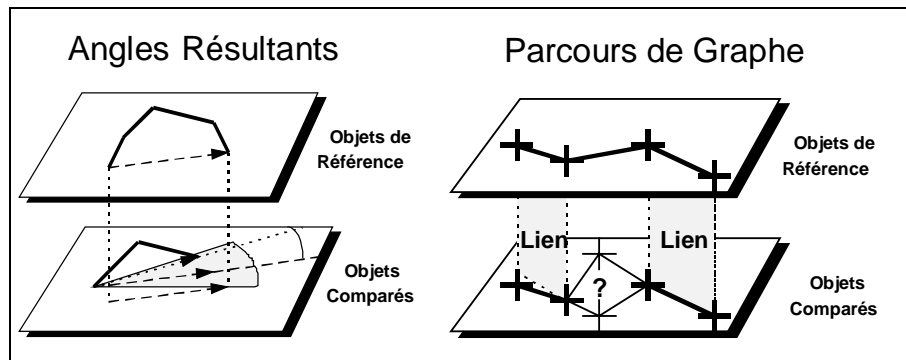


Figure 6. Méthodes employées lors de l'appariement des primitives linéaires

Afin de compléter et de valider les appariements des primitives ponctuelles et linéaires précédemment obtenus, une étape de regroupement et de vérification de ceux-ci est ensuite mise en œuvre. Un traitement particulier pour les carrefours complexes est également opéré. La description de ces deux processus fait l'objet de la section suivante.

2.1.3. Confrontation des appariements linéaires et ponctuels

Une fois l'appariement linéaire réalisé, on regroupe les appariements ponctuels en appariements n-m (i.e. : n objets de la base de référence correspondent à m objets de la base à comparer) afin d'identifier les intersections qui ont donné lieu lors de la mise à jour à des simplifications ou au contraire à un plus grand niveau de détails. En effet, les algorithmes utilisés pour générer les appariements ponctuels ne génèrent que des appariements 1-1, 1-n et n-1.

Ensuite, on parcourt tous les appariements linéaires générés, on analyse les nœuds extrêmes des lignes appariées et on vérifie, lorsque la distance des nœuds entre deux lignes appariées est inférieure au seuil fixé et que la ligne n'est pas un tronçon court, que l'appariement ponctuel existe bien.

Si cet appariement n'existe pas :

- Si aucun des nœuds intervenant dans l'appariement n'est impliqué dans un autre appariement, l'appariement est créé.
- Sinon, on demande à l'utilisateur de valider l'appariement à générer.

Cette étape, inspirée du processus décrit dans [GAB 94], permet non seulement d'améliorer sensiblement l'appariement mais aussi de détecter des erreurs qui pourraient avoir eu lieu dans la phase d'appariement ponctuel automatique.

Ensuite, on détecte les carrefours qui ont été détaillés ou au contraire simplifiés, ils correspondent dans une base à un nœud et dans l'autre base à plusieurs nœuds, chacun étant associé à un objet géographique. L'appariement est alors complété en ajoutant les primitives linéaires composant le carrefour à l'appariement ponctuel généré. Ensuite, les appariements linéaires sont mis à jour en supprimant les primitives linéaires impliquées dans les carrefours des appariements linéaires. Ces primitives étant en général de faibles longueurs elles apparaissent la plupart du temps dans des appariements linéaires parasites.

L'appariement des primitives ponctuelles et linéaires terminé, on passe à celui des objets géométriques surfaciques.

2.1.4. Appariement des primitives surfaciques

Ce type d'appariement est totalement indépendant des appariements ponctuel et linéaire. Il fait intervenir un certain nombre d'algorithmes basés sur des calculs d'aires afin de prendre en compte l'aspect surfacique des entités géographiques traitées. En effet, des méthodes d'appariement basées sur les contours font intervenir de façon plus prépondérante la forme des objets, et vont donc rejeter un grand nombre d'appariements pourtant valides.

Ainsi, afin de déterminer l'ensemble des entités surfaciques candidates à l'appariement, on utilise tout d'abord la notion de probabilité d'association définie dans [LEM 96a], [BEL 97], [VAU 98] et donnée ci-dessous (cf. figure 7) :

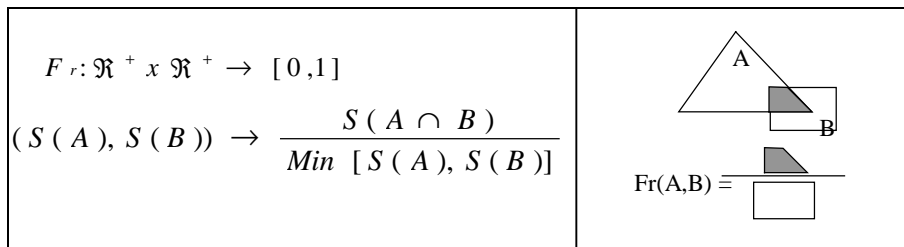


Figure 7. Probabilité d'association pour les entités surfaciques

Comme lors de l'appariement des primitives ponctuelles et linéaires, ce processus de filtrage implique le choix d'un seuil de rejet, qui dépendra des échelles et de la connaissance des indicateurs de qualité sur les jeux de données à traiter.

On utilise ensuite, la notion de distance surfacique [LEM 96a], [BEL 97], [VAU 98] définie comme suit (cf. figure 8) afin de générer à partir de l'ensemble des éléments candidats, les appariements surfaciques définitifs.

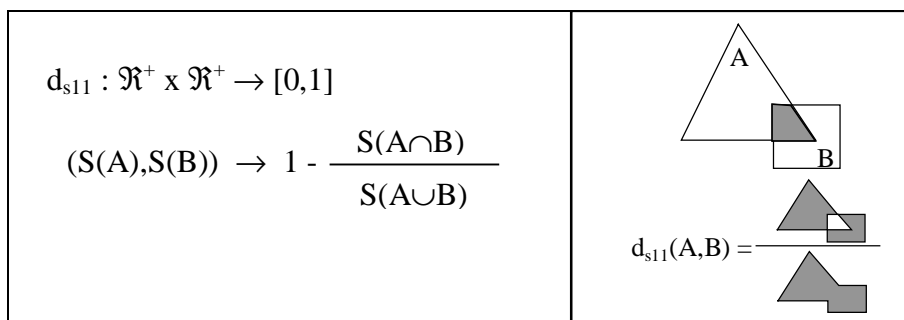


Figure 8. Définition de la distance surfacique

Il est à noter, que cette distance est définie entre deux entités surfaciques uniquement et implique également le choix d'un paramètre de rejet. Sa définition peut cependant être étendue à un ensemble de surfaces. En effet, A ou B peut être un agrégat d'entités surfaciques.

Cependant, cette dernière définition n'est pas utilisée lors de cette étape d'appariement. La détection des appariements 1-n est réalisée en parcourant la liste des appariements 1-1 établis. Ainsi, si une surface se retrouve impliquée dans plusieurs appariements 1-1, ceux-ci sont regroupés afin d'obtenir des appariements 1-n. On réalise donc une opération de détection des agrégats a posteriori. Cette méthode donne de bons résultats, mais un certain nombre d'appariements incomplets sont parfois produits et la remontée aux appariements n-m n'est pas toujours aisée, notamment lorsque les échelles des jeux de données sont très différentes. Ainsi, les agrégats ne sont pas tous détectés dans leur intégralité. Une idée d'amélioration de cette méthode serait de réaliser une recherche des agrégats a priori en utilisant les critères de groupement perceptif définis dans la théorie de Gestalt [ROC 83]. L'utilisation d'arbres de recouvrement minimum (« minimum spanning tree » en anglais) comme utilisés dans [REG 97] pour la généralisation du bâti permettrait ainsi de retrouver les agrégats avant toute opération d'appariement. La définition d'une distance entre agrégats serait alors à établir. Elle pourrait par exemple s'appuyer sur la distance entre les enveloppes convexes des différents groupements obtenus. Cette méthode est actuellement mise en œuvre au laboratoire COGIT de l'IGN et plusieurs distances sont actuellement testées.

L'appariement des primitives géométriques étant terminé, il faut maintenant procéder à celui des objets sémantiques. Le détail de cette étape est l'objet de la section suivante.

2.2. Appariement des objets simples et complexes

Une fois l'appariement des primitives réalisé, l'appariement des objets est très simple, il suffit pour chaque appariement de primitives de remonter aux objets qu'elles représentent. Or, il faut cependant ajouter des contraintes, en effet, sans contrainte sur les appariements possibles entre classe, le processus pourrait appairer des objets de natures très différentes. Il en résulterait alors des appariements incohérents entre objets sémantiques. Ainsi par exemple, lors de l'appariement de deux versions de la BDTopo®, un tronçon de route pourrait sans contraintes être apparié à un talus ou bien à une limite administrative ou encore à un tronçon hydrographique.

Ce mécanisme de contraintes permet donc outre de générer des appariements entre objets sémantiques qui soient cohérents, de gérer les changements de classe que peuvent subir certains objets lors notamment de changements de spécifications et/ou lorsque les jeux de données à traiter présentent des niveaux d'abstraction différents et/ou encore de retrouver des notions de partage de géométrie entre couches d'information.

Les contraintes d'appariement s'établissent donc d'un ensemble de classes à un autre ensemble de classes via l'utilisation d'un module qui analyse tous les appariements entre primitives et qui peuvent avoir lieu entre des classes différentes. L'utilisateur choisit alors interactivement pour chaque classe qui lui est proposée si oui ou non l'appariement est possible.

Les objets complexes quant à eux, ne disposant pas de géométrie, sont appariés en utilisant les appariements établis entre les objets simples les composants. Ainsi, deux objets complexes sont appariés si leurs composants le sont.

Enfin, il est à noter que ce processus d'appariement s'appuie uniquement sur les niveaux géométriques et topologiques de l'information géographique contenue dans les bases de données. La sémantique n'est à aucun moment utilisée. Cela permet de ce fait de réaliser un appariement tout à fait générique entre des jeux de données présentant des échelles, des niveaux d'abstraction et/ou de représentation très différents. Cela permet également de s'absoudre de la qualité sémantique parfois médiocre des bases de données mises en jeu. De même, lors de la recherche des mises à jour entre deux actualités d'une même base de données, ce mécanisme ne nécessite en aucun cas l'existence d'un système fiable et stable d'identifiants.

C'est d'ailleurs ce dernier point, c'est-à-dire la détection des évolutions spatio-temporelles subies par les entités géographiques entre deux actualités d'un même jeu de données que nous allons détailler dans la section suivante. Ainsi, nous verrons notamment comment passer des appariements générés entre objets, qui rappelons le sont des liens de correspondance entre ensemble d'entités représentant les mêmes phénomènes du monde réel dans deux représentations de celui-ci, aux mises à jour effectivement subies par celles-ci.

3. De l'interprétation des appariements à l'extraction des mises à jour

Afin de pouvoir catégoriser les mises à jour intervenues, ce mécanisme nécessite au préalable la définition d'une sémantique des évolutions spatio-temporelles que peuvent subir les entités géographiques. Ceci fait l'objet de la section suivante.

3.1. De la sémantique et de la logique des évolutions spatio-temporelles

Un certain nombre de typologies des évolutions spatio-temporelles que peut subir une entité géographique ont déjà été dressées dans la littérature. Elles présentent pour la plupart des approches différentes. En effet, certaines adoptent un point de vue proche de l'implantation de la notion de temporalité dans les systèmes d'information géographique tandis que d'autres tendent à traduire des évolutions plus proches de la réalité terrain, c'est-à-dire plus proche de phénomènes observables.

Ainsi, par exemple, dans [BED 97], seules les opérations de création et de destruction d'une entité géographique sont considérées. En effet, toute modification d'un objet de la base de données peut se traduire par la succession d'une opération de destruction de l'ancien objet et de création de la nouvelle entité. Cette typologie minimale a été volontairement retenue car il s'agissait dans le cadre de cette étude de structurer de façon simple et efficace l'information de mise à jour lors de sa livraison, afin d'en faciliter l'intégration. Il est à noter cependant que l'on perd ici le lien avec la nature réelle des évolutions spatiales subies par les entités géographiques modélisées dans la base de données.

Une autre approche intéressante est celle décrite dans [CHE 94]. Quatre grands types d'entités spatiales y sont définis : les entités spatiales aux contours fixes, les entités spatiales modifiables (i.e. : une entité à contour fixe formant une partition dynamique de l'espace, l'ensemble des entités la constituant pouvant varier de forme et d'attributs), les entités spatiales déformables (i.e. : entité formée d'un ensemble d'entités localisées dont les valeurs d'attributs, la forme et l'étendue peuvent varier) et les entités spatiales transformables (i.e. : dont la position et la forme évoluent au cours du temps). Cette typologie s'intéresse donc à l'identification des évolutions spatiales et attributaires subies par un ensemble d'entités géographiques. Si elles traduisent bien la réalité de phénomènes géographiques observables, la mise en œuvre d'une telle typologie dans l'extraction des mises à jour intervenues entre deux actualités d'une même base de données semble difficile.

Ainsi, l'approche exposée dans [CLA 94], définit une typologie des évolutions spatio-temporelles riche et plus facile à mettre en œuvre dans un système d'information géographique. Fidèle à l'analyse réalisée dans [GAG 93], celle-ci se décompose en un ensemble de processus de base, de processus de transformations et de mouvements applicables à une entité géographique. Les processus de base sont l'apparition, la disparition et la stabilité. Les processus de transformation sont quant à eux composés de l'expansion, de la contraction et de la déformation. Les

mouvements, enfin, sont soit le déplacement, soit la rotation. Cette typologie permet donc de rendre compte, de façon très détaillée, des évolutions spatio-temporelles que peut subir une entité géographique. Cependant, elle ne permet pas de transcrire facilement certaines modifications subies par un groupe d'entités spatiales ou par certains types d'objets géographiques modélisés dans les bases de données.

Ainsi, par exemple, la transformation d'un carrefour simple en un giratoire (cf. figure 9, gauche) semble difficilement rentrer dans cette typologie, si ce n'est par le biais de la succession d'une disparition et d'une apparition. Les notions de scission, de fusion, et d'agrégation (cf. figure 9, milieu et droite) tant pour les entités géographiques modélisées de façon linéaire (routier, hydrographie, ...) ou surfacique (bâti, parcellaire, ...) dans les bases de données semblent également difficiles à traduire dans les termes de cette typologie. Les modifications attributaires ne sont quant à elles pas prises en compte.

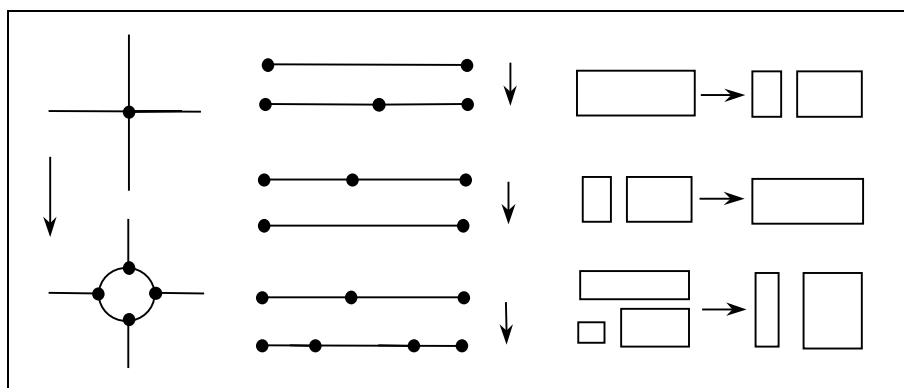


Figure 9. Exemples d'évolutions spatio-temporelles

Ceci nous pousse donc à établir une nouvelle typologie des évolutions spatio-temporelles que peuvent subir les entités géographiques afin de rendre compte au mieux des modifications intervenues entre deux actualités d'une base de données. Cette typologie, inspirée de celle proposée dans [CLA 94], est volontairement proche dans sa formulation des évolutions constatées dans les bases de données géographiques. Le but est, en effet, de donner à un utilisateur, une information détaillée sur la nature des évolutions intervenues afin de faciliter l'intégration des mises à jour dans son système d'information. Elle peut se décomposer en 8 grands types :

- La création (identique à l'apparition).
- La destruction (identique à la disparition).
- La scission (cf. figure 9, milieu et droite).
- La fusion (cf. figure 9, milieu et droite).
- L'agrégation (cf. figure 9, milieu et droite).
- La modification géométrique (regroupe la déformation et le déplacement).
- La modification sémantique.
- La stabilité.

Cette typologie permet donc de traduire les évolutions subies par une ou un ensemble d'entités spatiales et de prendre en compte tous les types d'objets géographiques, ainsi que les modifications attributaires.

L'utilisation de cette typologie dans un mécanisme d'extraction des mises à jour entre deux actualités d'une même base de données sera l'objet de la section suivante.

3.2. Le mécanisme d'extraction des mises à jour

Afin d'extraire les entités géographiques ayant subies des évolutions entre ces deux actualités et de classer ces changements dans les différents types précédemment définis, il faut analyser tous les appariements établis entre les deux jeux de données. L'analyse va notamment porter sur l'étude des cardinalités des liens d'appariement établis entre la base de données de référence et la base de comparaison, afin d'en déduire tout d'abord les évolutions purement géométriques et d'extraire ensuite les différences sémantiques entre objets qui n'ont pas forcément entraîné de modifications de la géométrie de ceux-ci. La base de référence est ici la base de données dans sa version la plus actuelle. La base de comparaison est alors la base de données dans sa version précédente.

Cette analyse, qui constitue le cœur du mécanisme d'extraction des mises à jour, peut se décomposer comme suit (cf. figure 10) :

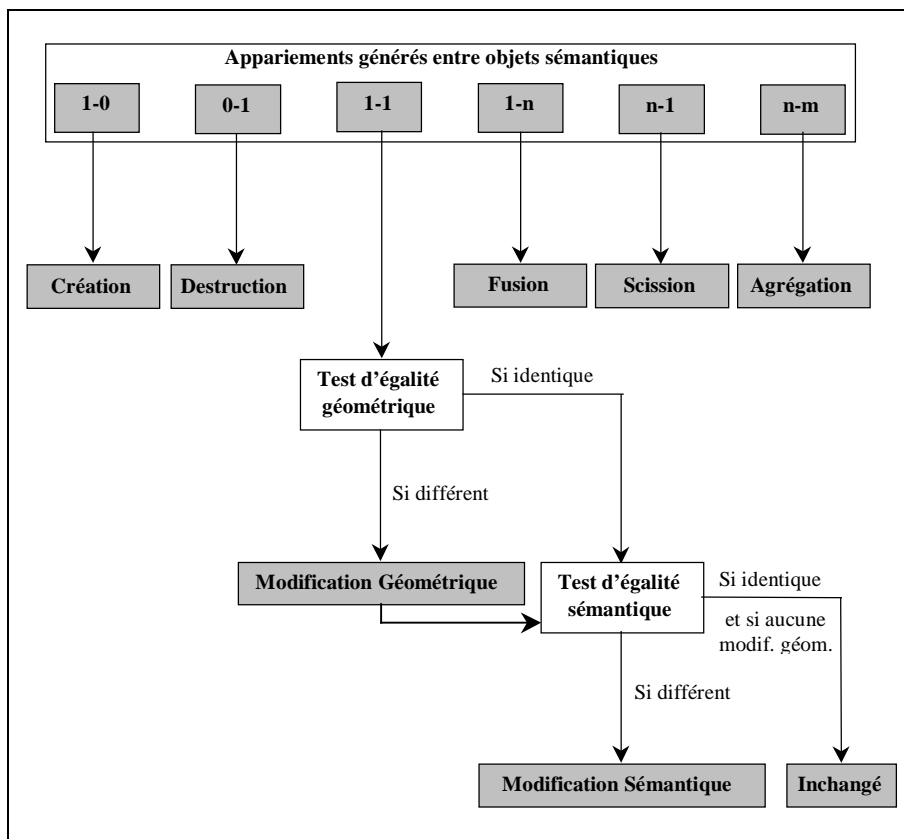


Figure 10. Fonctionnement du mécanisme d'extraction des mises à jour

Ainsi, les objets isolés (i.e. provenant d'appariements 1-0) du jeu de données de référence seront les objets géographiques créés.

Les objets isolés (i.e. provenant d'appariements 0-1) du jeu de données de comparaison seront considérés comme détruits.

Les objets intervenant dans des appariements 1-n (avec $n > 1$) seront considérés comme issus ou ayant subi une opération de fusion.

Les objets intervenant dans des appariements n-1 (avec $n > 1$) seront considérés comme issus ou ayant subi une opération de scission.

Les objets intervenant dans des appariements n-m (avec $n > 1$ et $m > 1$) seront considérés comme issus ou ayant subi une opération d'agrégation.

Les objets intervenant dans des appariements 1:1 sont analysés plus finement. En effet, on réalise tout d'abord un test d'égalité géométrique entre les entités. Si leurs géométries diffèrent, les objets seront considérés comme issus ou ayant subis une modification géométrique. Ensuite, les objets, qu'ils aient subi ou non une modification géométrique, sont comparés attributs par attributs car il se peut que des objets n'aient pas changé de géométrie mais seulement de contenu sémantique ou les deux. Si une différence est détectée, ils seront alors considérés comme issus ou

ayant subi une modification sémantique. C'est donc à ce stade seulement que le niveau sémantique de l'information géographique stockée dans les bases de données mises en jeu est pris en compte.

Enfin, les objets restants seront alors considérés comme stables ou inchangés.

Ce module d'extraction des mises à jour permet donc d'extraire non seulement des évolutions simples comme la création ou la destruction mais aussi de traduire des évolutions mixtes. En effet, les objets ayant changés à la fois de géométrie et de sémantique sont détectés et sont typés comme tels. Les opérations de fusion, de scission et d'agrégation ne sont pas quant à elles analysées plus finement, c'est-à-dire qu'on ne teste pas les évolutions de sémantique éventuelles que les entités ont pu subir. En effet, ces opérations sont en général dues à des changements sémantiques, le découpage ou le recollement de géométrie permettant ainsi de traduire l'homogénéité des attributs sur les nouveaux éléments géométriques définis.

Il est à noter également, que du fait des différents seuils fixés lors des étapes d'appariement, on peut ne pas accéder à la réelle nature de certaines évolutions subies par les entités modélisées dans les bases de données. En effet, un objet par exemple, peut se retrouver isolé dans l'une des deux bases car il a été rejeté lors d'une phase de l'appariement, du fait d'un choix de seuil trop bas. Il sera donc considéré comme créé ou détruit et non pas comme ayant subi une modification géométrique même si son homologue existe pourtant dans l'autre base de données. Il en va de même pour les opérations de fusion, scission ou agrégation, par exemple pour les éléments de réseaux routiers. En effet, des arcs symbolisant les tronçons de chaussée peuvent se retrouver impliqués dans les opérations précédemment citées sans pour autant partager le même support. Ceci est lié au fait que l'on autorise un certain écart de forme et de position, via le seuil fixé lors de l'appariement linéaire en terme de distance de Hausdorff (cf. figure 5). Le choix des différents seuils est donc à faire de façon judicieuse, si l'on veut accéder à une classification des mises à jour proche des évolutions réellement subies par les entités géographiques. Cependant, cela n'est pas rédhibitoire car ce processus traite les données dans leur intégralité. Ainsi, même si la nature de l'évolution n'est pas exactement celle intervenue, le changement sera tout de même détecté. Ce processus est donc complet, c'est-à-dire qu'il ne laisse échapper aucune modification des jeux de données, seule l'appartenance à un type d'évolution varie en fonction des seuils retenus lors des différentes étapes d'appariement. Les tests d'extraction des mises à jour menés au laboratoire COGIT de l'IGN, sur différents jeux de données, tendent d'ailleurs à prouver que le choix d'un seuil légèrement inférieur à l'erreur quadratique moyenne sur la position et la forme des objets constitue, dans ce cas, un choix tout à fait acceptable. La détermination automatique de ces seuils en fonction des bases de données mises en jeu est d'ailleurs l'objet d'une partie d'une recherche récemment engagée dans ce même laboratoire, l'idée étant de rendre le processus d'appariement totalement adaptatif.

Ainsi, ce mécanisme de détection des mises à jour entre deux actualités d'une même base de données géographiques, s'appuyant sur des méthodes d'appariement géométrique et topologique, permet donc d'extraire automatiquement les évolutions

subies par les entités spatiales modélisées dans les bases. Il ne nécessite en aucun cas l'utilisation ni même l'existence d'un système fiable et stable d'identifiants ou d'une gestion de l'historique dans les bases mises en jeu, et peut s'appliquer à tous types d'entités géographiques, à la différence des travaux réalisés dans [UIT 98], ne permettant de détecter les évolutions que du bâti. C'est donc à ce titre que le mécanisme présenté ici, peut être qualifié de générique.

D'autre part, l'information de mise à jour délivrée par ce module est détaillée et est proche des évolutions intervenant dans les bases de données géographiques. Il peut donc permettre à un utilisateur d'extraire, afin de les intégrer dans son système d'information, les évolutions d'un jeu de données de référence fourni par un producteur sous la forme d'une base entière mise à jour. Les risques de perte d'information ou d'obtention d'états incohérents causés par une intégration totale du nouveau jeu de données sont ici minimisés car l'utilisateur peut choisir clairement les modifications à répercuter ou non dans son système, sans se retrouver « noyé » dans le volume d'information délivré par le producteur et n'ayant pas forcément évolué.

Ce dernier point met donc en exergue les problèmes posés par la livraison de l'information de mise à jour pour les bases de données géographiques. En effet, à l'heure actuelle, peu de modes différents de livraison de cette information d'évolution sont mis en œuvre. Ainsi, la base entière mise à jour est-elle le plus souvent fournie à l'utilisateur. Aussi allons-nous, dans la section suivante, nous attacher à voir en quoi ce module de détection des mises à jour peut nous permettre de définir de nouveaux modes de livraison et notamment la notion de calques différentiels de mise à jour.

4. Vers la livraison de calques différentiels de mise à jour

Comme nous l'avons mentionné précédemment, peu de modes différents de livraison des mises à jour dans les bases de données géographiques sont actuellement utilisés. Cela tient en grande partie à la spécificité de l'information manipulée. En effet, s'il est relativement aisé de traduire dans le cadre des bases de données dites « classiques », les modifications attributaires subies par les objets, il n'en va pas de même pour l'information géométrique stockée dans les bases de données géographiques. Cependant, les notions de différentiels d'états et de journaux pourraient être assez facilement transposés afin de délivrer les mises à jour de façon détaillée et minimale.

Les différentiels d'état permettent ainsi de délivrer les changements sous la forme de couples d'objets anciens/nouveaux ou plus simplement sous la forme de références sur les objets anciens/objets nouveaux, si un système d'identifiants est bien entendu mis en œuvre dans la base. La livraison des mises à jour par journaux n'implique quant à elle aucune livraison d'objets, seule la description des transformations permettant de passer de l'ancien état des objets (repérés par leur identifiant) au nouveau est délivrée.

Ainsi, la structuration de l'information fournie par le module de détection et

d'extraction des mises à jour, présenté précédemment, semble relativement proche de celles contenues dans les différentiels d'état et les journaux. En effet, du fait de l'établissement des liens d'appariement entre entités, nous disposons, en quelque sorte, de la « filiation temporelle » des objets. Les mises à jour étant effectivement déduites des appariements, on connaît quel objet ou groupement d'objets succède à une entité ou à un groupe d'entités modélisées dans les différentes versions de la base de données. On dispose également de la nature de l'évolution subie.

De ce fait, un mode permettant de livrer l'information de mise à jour de façon détaillée et minimale pourrait être de fournir la géométrie des objets dans leur état ancien et nouveau, ainsi que la nature des évolutions subies par chaque entité au travers d'un journal. C'est ce que nous nommons les calques différentiels de mise à jour.

Afin d'illustrer ce mode de livraison, des résultats issus d'un test du module de détection des mises à jour, effectué sur le réseau routier du département du Calvados (14) de la BDCarto®, vont maintenant être présentés.

Ainsi, le calque fournit ci-après (cf. figure 11) donne une représentation de la géométrie des objets issus d'une évolution, c'est-à-dire contenus dans la version à jour de la BDCarto®. Il est appelé « calque différentiel +1 ».

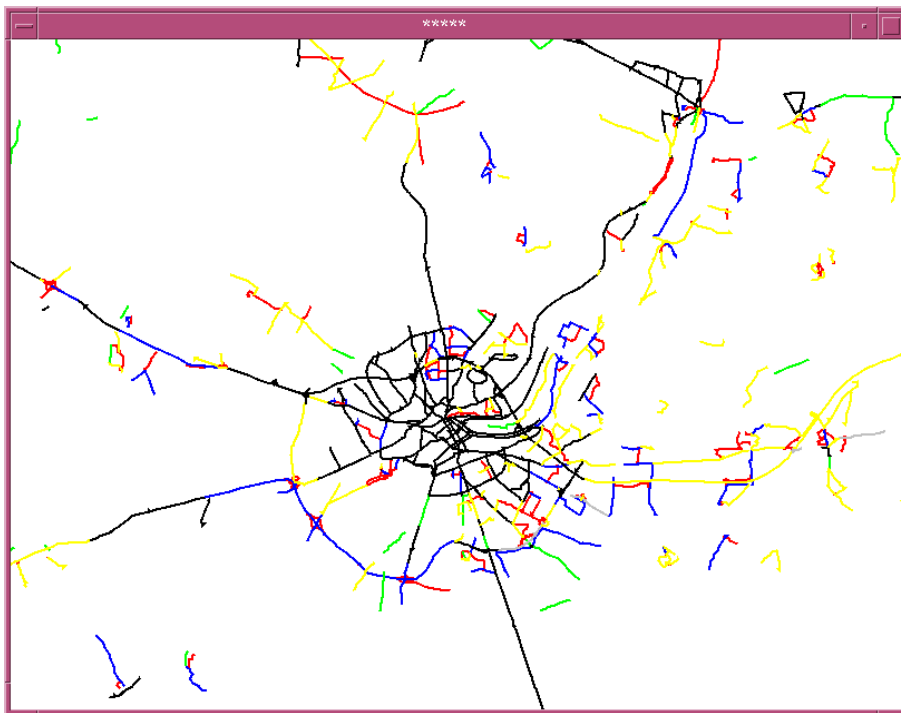


Figure 11. Extrait du calque différentiel +1 de la BDCarto® (Caen, Calvados)

Les différentes couleurs présentes sur ce calque traduisent la nature des évolutions dont sont issues les entités. Ainsi, les objets créés, ou issus d'une fusion, d'une scission, d'une agrégation, ou d'une modification géométrique et/ou sémantique sont représentés.

De la même façon, le calque donné ci-après (cf. figure 12) donne une représentation de la géométrie des objets ayant subi une évolution, c'est-à-dire contenus dans la version périmée de la BDCarto®. Il est appelé, par analogie avec le précédent, « calque différentiel -1 ».

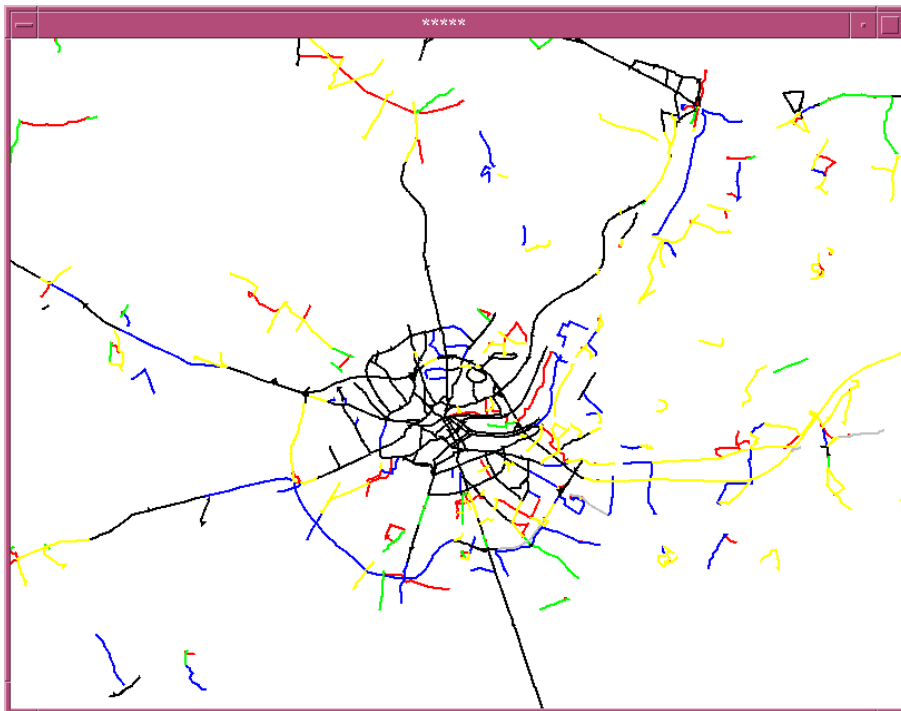


Figure 12. Extrait du calque différentiel -1 de la BDCarto® (Caen, Calvados)

De même que précédemment, les différentes couleurs présentes sur ce calque traduisent la nature des évolutions subies par les entités. Cette fois-ci, les objets détruits, ou ayant subis une fusion, une scission, une agrégation, ou une modification géométrique et/ou sémantique sont représentés.

La nature des évolutions subies par chaque entité, ainsi que le détail des évolutions sémantiques sont traduits dans l'extrait de journal suivant (cf. figure 13) :

Journal des mises a jour : BDCarto v1 -> BDCarto v2

Créations:

NEUD_ROU -> BDC_ID: 991584043
 TRON_ROU -> BDC_ID: 992300803
 TRON_BAC -> BDC_ID: 269
 LIAISON -> BDC_ID: 153
 CARR_CX -> BDC_ID: 1375
 ROUTE -> BDC_ID: 39095
 ...

Destructions:

NEUD_ROU -> BDC_ID: 140005977
 TRON_ROU -> BDC_ID: 991363378
 ROUTE -> BDC_ID: 24254
 ...

Fusions:

TRON_ROU : 140004576 TRON_ROU : 140004577 -> TRON_ROU : 140004577
 TRON_ROU : 990157858 TRON_ROU : 140006846 -> TRON_ROU : 992260352
 ...

Scissions:

TRON_ROU : 140004259 -> TRON_ROU : 992260418 TRON_ROU : 992260417
 TRON_ROU : 140002569 -> TRON_ROU : 992301214 TRON_ROU : 992300918
 TRON_ROU : 991744402 -> TRON_ROU : 992300352 TRON_ROU : 992300349 TRON_ROU :
 992300351
 ...

Agrégations:

TRON_ROU : 140003594 TRON_ROU : 990746918 -> TRON_ROU : 992300653 TRON_ROU :
 992301195
 TRON_ROU : 991520666 TRON_ROU : 991520657 -> TRON_ROU : 992301382 TRON_ROU :
 992301381 TRON_ROU : 992301376
 ...

Modifications Sémantiques:

NEUD_ROU : 990039893 -> NEUD_ROU : 990039893 , TYPE: 4 -> 5 COTE: 0 -> 9999
 TRON_ROU : 140004360 -> TRON_ROU : 140004360 , NBV_TOT: 1 -> 3
 ...

Modifications Géométriques:

NEUD_ROU : 140000215 -> NEUD_ROU : 991597654
 TRON_ROU : 990747229 -> TRON_ROU : 992301035
 ...

Figure 13. Extrait du journal des évolutions de la BDCarto® (Calvados)

Ce journal s'appuie donc, dans sa formulation, sur le système d'identifiants présent dans la BDCarto®. Cependant, on peut imaginer assez facilement des structures permettant de délivrer la même information, sans l'existence d'un tel système dans les bases de données mises en jeu. Ainsi, par exemple, on pourrait livrer pour chaque évolution, la liste des géométries des objets sous la forme de couple anciennes/nouvelles, avec pour chaque couple, la liste des modifications sémantiques, s'il y en a bien entendu, avec le nom de l'attribut concerné et le type d'opération de mise à jour qu'ont subi les entités géographiques, conformément à la typologie définie précédemment. Un système d'identifiants temporaires, générés à la volée, permettant de référencer les objets au sein de l'échange et uniquement au sein de celui-ci, pourrait également être satisfaisant.

Ce mode de livraison de l'information de mise à jour permet donc de fournir à un utilisateur, une description minimale (en terme de nombre d'objets) mais détaillée des évolutions qui ont eu lieu dans le jeu de données de référence. Il pourra ainsi, maîtriser plus facilement l'intégration de ces mises à jour dans son propre système d'information, en choisissant par exemple quelles modifications répercuter ou non. Il est à noter que ce mode pourrait être combiné avec la livraison de la base entière ancienne ou à jour afin de retrouver des objets malencontreusement détruits lors d'une manipulation ou bien pour reconstituer l'information de contexte avec les voisins de l'objet ayant évolué.

Cette notion de contexte est d'ailleurs relativement importante car elle pose indirectement le problème du format de l'échange. En effet, tous les formats ne permettent pas de délivrer des différentiels ou alors pas de manière simple. Par exemple, EDIGÉO [AFN 92] ne permet de le faire qu'en se plaçant dans un niveau non topologique et en livrant des géométries fictives, ce qui peut représenter un danger pour un utilisateur non averti. Il est à noter que des réflexions quant au choix de la forme et du format de livraison des différentiels sont en cours à l'IGN pour la livraison notamment des mises à jour de la BDCarto®.

5. Conclusion

Ainsi, ce mécanisme de détection des mises à jour entre deux actualités d'une même base de données géographiques, s'appuyant sur des méthodes d'appariement géométrique et topologique, permet donc d'extraire automatiquement les évolutions subies par les entités spatiales modélisées dans les bases. Il ne nécessite en aucun cas l'utilisation ni même l'existence d'un système fiable et stable d'identifiants ou d'une gestion de l'historique dans les bases mises en jeu. C'est à ce titre qu'il peut être qualifié de générique.

D'autre part, l'information de mise à jour délivrée par ce module est détaillée et est proche des évolutions intervenant dans les bases de données géographiques. Il peut donc permettre à un utilisateur d'extraire, afin de les intégrer dans son système d'information, les évolutions d'un jeu de données de référence fourni par un producteur sous la forme d'une base entière mise à jour. Les risques de perte

d'information ou d'obtention d'états incohérents causés par une intégration totale du nouveau jeu de données sont ici minimisés car l'utilisateur peut choisir clairement les modifications à répercuter ou non dans son système, sans se retrouver « noyé » dans le volume d'information délivré par le producteur et n'ayant pas forcément évolué.

La possibilité d'utiliser ce mécanisme dans la définition de nouveaux modes de livraison de l'information de mise à jour a également été détaillée. La notion de calques différentiels de mise à jour a notamment été définie et un premier exemple de mise en œuvre a été donné. Ils constituent donc des premières aides à l'intégration, dans les bases de données utilisateur, des mises à jour fournies par le producteur.

Il est enfin à noter que la définition de ce module s'intègre du fait de ses propriétés dans une étude plus large, menée actuellement au laboratoire COGIT de l'IGN, visant à obtenir un outil de mise à jour générique, le plus automatique possible pour les bases de données multi-représentations. La structure de ce mécanisme plus général peut notamment être trouvée dans [BAD 98].

6. Remerciements

L'auteur tient ici à remercier tout particulièrement Cécile Lemarié, qui travaille depuis plusieurs années sur l'appariement de données géographiques au laboratoire COGIT de l'IGN et qui a de ce fait contribué à l'élaboration de ce travail.

7. Bibliographie

- [ABB 94] I. Abbas. Base de données vectorielles et erreur cartographique : problèmes posés par le contrôle ponctuel - Une méthode alternative fondée sur la distance de Hausdorff : le contrôle linéaire. Thèse de doctorat de l'université de Paris XII, Paris, France, 10 juin 1994.
- [AFN 92] AFNOR. Echange de données informatisées dans le cadre de l'information géographique (EDIGéO), AFNOR, 1992, ISSN 0335-3951.
- [BAD 98] T. Badard. Towards a generic updating tool for geographic databases. To appear in the proceedings of GIS/LIS'98, Annual Exposition and Conference. Fort Worth, Texas, USA, November 10-12, 1998, 12 pages.
- [BED 97] Y. Bédard, Y. van Chestein, G. Poupart-Lavoie. Actualisation des données à référence spatiale (volet échange et intégration). Centre de Recherche en Géomatique, Université Laval, Québec, Québec, Canada, 54 pages.
- [BEL 97] A. Bel Hadj Ali. Appariement géométrique des objets géographiques et étude des indicateurs de qualité. Mémoire de DEA Sciences de l'Information Géographique, Université de Marne la Vallée, Marne la Vallée, France, 18 Septembre 1997, 109 pages.
- [CHE 94] J. P. Cheylan, S. Lardon, H. Mathian, L. Sanders. Les problématiques liées au temps dans les SIG. Dans la Revue internationale de géomatique. Vol. 4, Numéro 3-4, 1994, pp. 287-305.

- [CLA 94] C. Claramunt, M. H. de Séde, R. Prélaz-Droux, L. Vidale. Sémantique et logique spatio-temporelles. Dans la Revue internationale de géomatique. Vol. 4, Numéro 2, 1994, pp. 165-180.
- [GAB 94] Y. Gabay, Y. Doytsher. Automatic adjustment of line maps. In the proceedings of GIS/LIS'94, Annual Conference and Exposition. Phoenix, Arizona, USA, October 25-27, 1994, pp. 333-341.
- [GAG 93] P. Gagnon. Concepts fondamentaux de la gestion du temps dans les SIG. Mémoire de Maîtrise, Université Laval, Québec, Québec, Canada, 1993, 136 pages.
- [LAU 92] R. Laurini, D. Thompson. Fundamentals of spatial information systems. Academic Press, London, 1992.
- [LEM 96a] C. Lemarié. Etat de l'art sur l'appariement. Rapport technique DT/960022/S-RAP, IGN, Service de la Recherche, Saint Mandé, France, juillet 1996, 37 pages.
- [LEM 96b] C. Lemarié, L. Raynal. Geographic data matching: First investigations for a generic tool. In the proceedings of GIS/LIS'96, Annual Conference and Exposition. Denver, Colorado, USA, November 19-21, 1996, pp. 405-420.
- [LEM 98] C. Lemarié, O. Bucaille. Spécifications d'un module générique d'appariement de données géographiques. Dans les actes du 11ème congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'98). Clermont-Ferrand, France, pp. 397-406.
- [LUP 87] A. E. Lupien, W. H. Moreland. A general approach to map conflation. In the proceedings of the ACSM/ASPRS Annual Convention & Exposition, Technical Papers, AutoCarto 8, Baltimore, Maryland, USA, March 29-April 3, 1987, pp. 630-639.
- [RAY 96] L. Raynal. Some elements for modelling updates in topographic databases. In the proceedings of GIS/LIS'96, Annual Exposition and Conference. Denver, Colorado, USA, November 19-21, 1996, pp. 1223-1232.
- [REG 97] N. Regnauld. Généralisation du bâti : Structure spatiale de type graphe et représentation cartographique. Mémoire de thèse en informatique de l'université de Marseille, Marseille, France, 1997, 179 pages.
- [ROC 83] I. Rock. The logic of perception. Cambridge, Massachusetts, MIT Press.
- [SAA 88] A. Saalfeld. Conflation – Automated map compilation. In the International Journal of Geographical Information Systems (IJGIS), Volume 2, Number 3, Taylor & Francis, July-September 1988, pp. 217-228.
- [UIT 98] H. Uitermark, P. van Oosterom, N. Mars, M. Molenaar. Propagating updates: Finding Corresponding objects in a multi-source environment. In the proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98), T. K. Poiker and N. Chrisman Eds., Vancouver, BC, Canada, July 11-15, 1998, pp. 580-591.
- [VAU 98] F. Vauglin, A. Bel Hadj Ali. Geometric matching of polygonal surfaces in GIS. ASPRS-RTI Annual Conference, Tampa, Floride, USA, March 30-April 3, 1998, pp. 1511-1516.

Using XML for the exchange of updating information between geographical information systems

Thierry Badard and Didier Richard
Institut Géographique National
Service de la Recherche/Laboratoire COGIT
2 à 4, avenue Pasteur
94165 – Saint Mandé Cedex, France
Email : {Thierry.Badard, Didier.Richard}@ign.fr

1 INTRODUCTION

Nowadays, Geographic Information Systems are considered to be truly analysis and decision-making tools. For that reason, a growing number of organisations invest in such systems and add specific information necessary to the tasks for which they have the responsibility. However, the implementation of such systems is difficult and relatively expensive. That is why these institutions purchase reference data sets from geographic information producers. From these they are able to develop their own information systems. Thus, these organisations are considered, by the producer, as users of reference geographic data sets.

To carry out their assignments, these institutions (e.g. governmental agencies—Department of Public Works, Environmental Agency—, public utilities—electricity, gas, water—, private companies with a national remit, etc.) clearly need updates from the producer, in order to have the most faithful and realistic image of geographic reality. However, at the present time, the integration of these updates is still problematic. A taxonomy of the problems hindering the integration of the updates for geographic databases has already been presented and detailed in (Badard, 1998). This taxonomy has especially pointed up that few different delivery modes of the updates are available for geographic databases. The whole up-to-date database is often delivered even if only 10% of the objects stored in the database have changed. This is a common estimation of the rate of evolution per year for geographic databases (Raynal, 1996). Users are thus compelled to integrate the whole new version of the database in their information systems, which may induce significant risks of information loss or leave the database in an inconsistent state. Otherwise, they have to retrieve the updates by themselves. This task may fail if a reliable and stable system of identifiers has not been implemented in the database. Other solutions have been proposed but they do not fulfil all the requirements for an easy and consistent integration of evolutions in geographical information systems. Besides, these solutions come up against interoperability issues between GIS software.

This paper focuses on the implementation of a solution for the exchange of updating information between geographical information systems, which has been developed at the COGIT laboratory of IGN (the French National Mapping Agency). It is based on XML

(eXtensible Markup Language) and aims at providing users with minimal, structured and detailed information concerning evolutions in order to make their integration in systems easier and more consistent. This solution allows the concerns raised by the lack of interoperability between GIS software to be tackled and to address the problem of real time updating for geographic databases in client/server architectures.

In section 2, challenges and benefits brought by XML for the transfer of evolutions in geographic databases are first discussed. The conceptual model defined and implemented at the IGN, which enables such a transfer is presented in section 3. Examples of implementation on data from BDCARTO® (a 1:100 000 geographic database with 10-meter resolution produced by the IGN) are provided and detailed in section 4.

2 WHY XML FOR THE DELIVERY OF UPDATING INFORMATION ?

To address the issue of the delivery of updating information for geographic databases, early solutions have been proposed but they do not fulfil all the requirements for an easy and consistent integration of the evolutions in the geographical information systems. For instance, Poupart-Lavoie (1997) has defined a method for the transfer of updates between the geographic databases used at the Ministry of Natural Resources in Quebec. This method is based on the relational technology and provides minimal and detailed information on changes. It is unfortunately difficult to implement in another context. Badard (1999) has defined the notion of “updating delta”. It consists in the minimal delivery of the new and old versions of the updated objects. The nature of the evolutions undergone by the geographical entities is then provided and detailed in a log. Due to the lack of interoperability between GIS software, this solution implies the definition of a specific information structuring for each format. Moreover, it often leads to the delivery of exchanges difficult to integrate because GIS data formats have not been originally designed for the transfer of updating information. For that reason, the International Hydrographic Organisation (1996) has designed a specific data exchange format for hydrographic data, which includes the delivery of evolutions. It allows real time updating of geographic databases, but is specifically devoted to the management of hydrographic data sets, which prevents its application for other geographic data.

Together with the development of these solutions, the computer science community has seen the emergence of XML, a new data format for the exchange of structured documents on the web. XML stands for eXtensible Markup Language and like HTML, it is a subset of SGML, the Standard Generalized Markup Language (ISO 8879). Version 1.0 of the XML specifications was accepted by the World Wide Web Consortium (W3C) as a recommendation on February 10, 1998 (World Wide Web Consortium, 1998), and is likely to become a widespread standard. XML aims at enabling generic SGML to be served, received, and processed on the web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML. While HTML is powerful for formatting a document, XML describes the document’s structure and meaning.

Although many applications of XML are Internet-related via the implementation of Java and web-based client/server data access architectures (as illustrated in Figure 1), XML is in no way limited to Internet use. In fact, XML is designed for organising information, which makes it perfect for exchanging data between different systems, regardless of whether the Internet is part of the picture. It thus addresses both archive purposes and deployment in situations where continuous high bandwidth connection-by-

wire solutions are inappropriate. XML specific compression techniques are being developed in order to resolve data volumes issues.

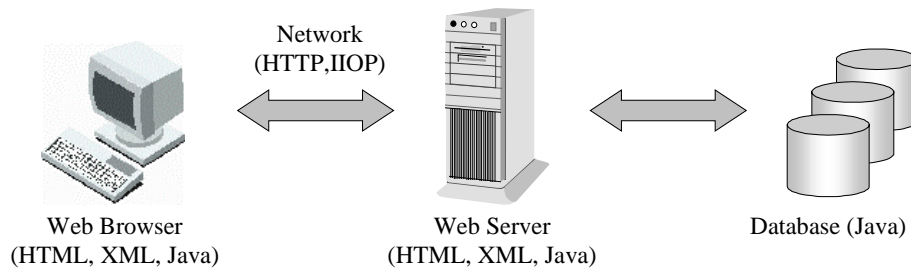


Figure 1 Java and Web-based Client/Server Data Access.

In addition, XML provides a mechanism, named DTD (Document Type Definition) for the design of self-describing data exchanges. It allows for the definition of domain specific markup languages by providing developers with capabilities to set up their own syntax (i.e. their own tags). A sample of a DTD designed for the modelling of geometrical features (points and lines) is presented in Figure 2.

```

<?xml version="1.0" encoding="UTF-8"?>
...
<!-- Definition of points (0 dimension) : -->
<!-- Co-ordinates are expressed in a planimetric reference system (RPLA code) -->
<!--and in an altimetric system (RALT code), if required. -->
<!ELEMENT AIG.Pt EMPTY>
<!ATTLIST AIG.Pt RPLA ENTITY #IMPLIED
      UPLA CDATA #IMPLIED
      RALT ENTITY #IMPLIED
      UALT CDATA #IMPLIED
      X CDATA #REQUIRED
      Y CDATA #REQUIRED
      Z CDATA #IMPLIED>
<!-- Definition of lines (1 dimension) : -->
<!ELEMENT AIG.Li (AIG.Pt)+>
<!ELEMENT AIG.LiRef EMPTY>
<!ATTLIST AIG.LiRef %AIG.idref;>
...

```

Figure 2 Example of DTD (Document Type Definition).

It thus enables the support of a wide variety of applications. It is not restricted to textual data and it is independent from the formats and the information structuring managed by the different communicating systems. If an application understands XML, it automatically understands all the languages built from it. Besides, XML provides the concepts of well-formedness and of validity, which ensure that an exchange conforms respectively to the XML specifications and to the syntax defined in the DTD.

Considering the challenges and the benefits of such a language, many GIS software companies plan to make their platforms XML compliant. Laser-Scan has for instance announced the release of SOTF (Spatial Object Transfer Format) on March 30, 2000. SOTF results from a research project in collaboration with NIMA (contract NMA201-98-3-0022). It is an XML encoding of geographic data designed to support incremental update, preservation of value-added data, client "area of interest" and schema transfer

with no reliance on tiling. Some interesting concepts have been introduced, but metadata and topology are for instance unfortunately not supported by SOTF.

The OpenGIS Consortium has started working on the specifications of an XML encoded transfer of geographical features for display purposes on web map servers. It has resulted in the release of two specifications documents: SFXML (Simple Feature XML) for the modelling of geometrical features (OpenGIS Consortium, 1999a) and GML (Geography Markup Language) for the implementation of geographical objects (OpenGIS Consortium, 1999b). As in SOTF, no notions of topology, geometry sharing, multiple-primitives, metadata and explicit relationships between geographical objects are supported by either SFXML or GML.

Finally, XML has been adopted for metadata transfer between institutions in Denmark. This transfer relies on the ISO/TC 211 standard for metadata, for which ISO has released a specific DTD (ISO/TC 211, 1999).

Considering the benefits provided by such a language, the use of XML has been investigated at IGN in order to tackle the issues that hinder the delivery of updating information for geographic databases. It results in the definition of a conceptual model, which enables the transfer of evolutions in such a language. This model is detailed in the next section.

3 A CONCEPTUAL MODEL ENABLING THE DELIVERY OF EVOLUTIONS

As mentioned in the previous section, this conceptual model allows for the delivery of detailed and structured updating information for geographic databases. The simplified structure of this conceptual model is illustrated in Figure 3. It can be decomposed into 4 main parts.

The first part concerns “metadata” (Figure 3, middle). It deals with the descriptive information related to the different geographic databases for which updates are delivered. It includes the schemas of the different databases and specifies for instance their spatial reference systems (direct and indirect), their extents, their co-ordinate units and their dates of capture or last update. This allows producers to supply users with data sets with different extents, co-ordinate unit systems or at different dates on request. This part of the model relies on the French profile of the European prestandard ENV 12657 (CEN/TC 287, 1999), which has been proposed by CEN/TC 287. This choice results from the national context in which the issue of the delivery of updating information for geographic databases has been addressed. The CEN/TC 287 prestandard is indeed commonly used in France for the management of metadata. Only what is defined as mandatory information in this prestandard has been implemented in the model. Information dealing with data quality for instance has not been taken into account, but could be easily added in the future.

The second part of the model deals with the “evolutions” (Figure 3, top). It structures and details the evolutions undergone by the geographical entities stored in the databases. Information modelled in this part relies on the taxonomy of evolutions drawn up in (Badard, 1999). The different types of evolution proposed in this taxonomy are deliberately detailed and close to the effective modifications performed in geographic databases in order to make their integration in the systems easier and more consistent. The available operations are thus creation, deletion, merging, splitting, aggregation, semantic and/or geometrical modification. These operations can concern not only the objects but also the schema of the databases. They have been implemented in the model.

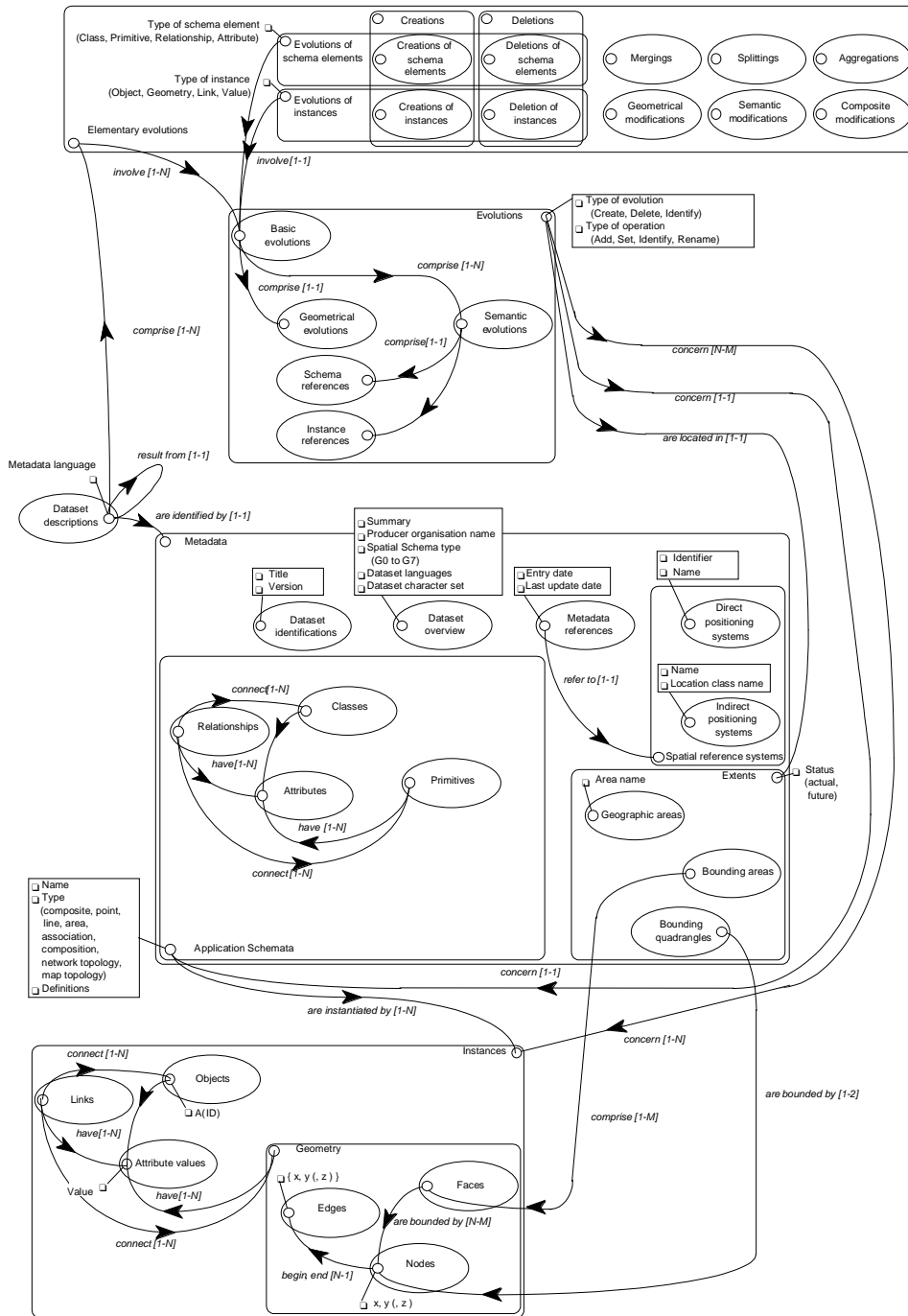


Figure 3 Conceptual model enabling the transfer of evolutions with XML, expressed in the HBDS formalism (Bouillé, 1987).

The third part of the model concerns the “geographical application” (Figure 3, bottom). In accordance with the metadata, this part allows for the modelling of the different elements modified in the database: objects, attributes and their values, relationships (between objects, between geometrical features, and between objects and geometrical features), and geometrical features. It supports topological relationships and geometrical features with multiple primitives and different representations. This part of the model is based on the European prestandard ENV 12160 (CEN/TC 287, 1997) established by CEN/TC 287 and enhances it. Concepts related to object-orientation and allowing an instance (object, attribute value, link, geometry) to be linked with its type (class, attribute, relationship, primitive) have for instance been added. It thus enables the delivery of the modifications undergone by the schemas of the databases. This part of the model also relies on the identification principle: every element of the database has an identifier. The identifier value is computed with the MD5 algorithm (Network Working Group, 1992). This algorithm often implemented on web servers to identify connections, enables the calculation of a unique 128-bit encrypted identifier, independently from hardware platforms. The use of this algorithm in the geographic context allows for the delivery of updating information even if no system of identifiers has been implemented in any of the geographic databases: users have just to consistently identify their data sets (i.e. calculate the identifiers with the relevant input data: attribute values, geometry and/or contextual information) before processing the updates delivered in the exchange.

The final part of model enables the “evolution transfer” (Figure 3, middle left). It allows for the consistent delivery of the evolutions by linking together the various components of the exchange. It provides users with information dealing with the number of transferred updates and the location of the descriptions of these evolutions in the exchange.

The use of XML enables thus the transfer of both minimal and detailed information concerning the evolutions: A first message (in XML) will warn the users that one or several evolutions will be transferred. Messages concerning metadata and evolutions for each version of the database will then follow, and messages relating to the geographical application will finally be sent.

In order to validate the previous model and prove its feasibility, a set of DTD and a prototype have been implemented at the COGIT laboratory of IGN. A test has also been performed with the road network of BDCARTO@ (a 1:100 000 geographic database with 10-meter resolution produced by the IGN). In order to illustrate this process, examples of the different XML messages generated in this test by the prototype are presented and detailed in the next section.

4 IMPLEMENTATION OF THE DIFFERENT XML MESSAGES

The prototype has been implemented in GeO₂ (Raynal *et al.*, 1995), an objected-oriented database management system for geographic databases developed at the COGIT laboratory of IGN. Examples of the different XML files generated by the prototype are provided and detailed below. Except for tags, each message has been anglicised. These messages concern a semantic and geometrical modification (named composite modification). Each message is well-formed and valid. Although XML is not merely devoted to Internet, comments provided in this section are deliberately related to client/server architectures in order to understand how a system of real time updating for geographic databases could process information delivered on a network: a server (producer) provides thus one or several client systems (users) with updating information encoded in XML.

As mentioned in the previous section, users' systems are warned that some updates are incoming with the delivery of an "evolution transfer" message. It allows the consistent delivery of the updating information by linking the different components of the exchange. Figure 4 illustrates the content of such a message.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- trloevdep22.xml -->
<!DOCTYPE TrEvolInfoGeo SYSTEM "TransfEvol.dtd">
<TrEvolInfoGeo>
  <TIG.LotEvol>
    <TIG.Reference_lot_ancien xlink:href="dep22bdc20_cta_1994.xml#ROUTES94" />
    <TIG.Reference_lot xlink:href="dep22bdc20_cta_1998.xml#ROUTES98" />
    <TIG.Evolution xml:title="List of the evolutions for the road network theme.">
      <TIG.EvolRef xlink:href="dep22evol1.xml#_1" />
      <TIG.EvolRef xlink:href="dep22evol2.xml#_2" />
    ...
  </TIG.Evolution>
  <TIG.ApplGeo xlink:href="dep22rte.xml" />
</TIG.LotEvol>
</TrEvolInfoGeo>
<!-- End of the evolution transfer trloevdep22.xml -->
```

Figure 4 Sample of an "evolution transfer" message (trloevdep22.xml).

The first line of the file informs users that this XML message is encoded with the ISO-8859-1 character set and according to version 1.0 of XML specifications. The DOCTYPE tag mentions that the content of this file relies on a specific DTD: Syntax (i.e. the different tags) used in this message to structure information is therefore defined in a separate file (TransfEvol.dtd).

After these header lines necessary to make the users' systems able to understand the content of the message, references to metadata concerning the two versions of the geographic database are provided. They concern the road network themes ROUTES94 and ROUTES98, respectively described in the dep22bdc20_cta_1994.xml and dep22bdc20_cta_1999.xml files. The evolutions undergone by the geographical objects of the road network are then listed. Each evolution is identified and detailed in a specific XML message. For instance, "dep22evol2.xml#_2" corresponds to the second evolution transferred to the users' systems. Its description can be found in the dep22evol2.xml file. Finally, a reference to the file describing the "geographical application" part of the exchange is provided (dep22rte.xml).

Once this first message has been processed by the system, files describing metadata of the two versions of the geographic database are delivered to the users' systems. Figure 5 provides a sample of such an XML file. It corresponds to the metadata of the up-to-date version of the database.

```
<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- dep22bdc20_cta_1998.xml -->
<!DOCTYPE Metadonnees.Description SYSTEM "XP-ENV12160.dtd">
<Metadonnees.Description>
  <Metadonnees.Lot.Identification titre="ROUTES98" version="BDCA_CTA">
  </Metadonnees.Lot.Identification>
  <Metadonnees.Lot.Apercu nom_du_producteur="IGN">
    <resume>No comments</resume>
    <topologie type="spatialG1"></topologie>
    <langue />
  </Metadonnees.Lot.Apercu>
  <Metadonnees.Lot.Reference ref_spatial_plani="BP:LAMBE">
```

```

<date_saisie>1998-04-07</date_saisie>
<date_derniere_maj>1998-06-05</date_derniere_maj>
</Metadonnees.Lot.Reference>
<Metadonnees.SystemeSpatial>
  <Positionnement.Direct acronyme="BP:LAMBE" unite-plani="cm" />
</Metadonnees.SystemeSpatial>
<Metadonnees.Lot.Emprise>
  <Emprise.actualite>
    <date_debut>1998-06-05</date_debut>
    <date_fin>1999-12-01</date_fin>
  </Emprise.actualite>
  <Emprise.planaire>
    <quadrilatere>
      <min x="15767276" y="234831998" />
      <max x="28706101" y="244424901" />
    </quadrilatere>
  </Emprise.planaire>
</Metadonnees.Lot.Emprise>
<Metadonnees.Lot.Definition>
  <description_generale>
    Road Network of BDCarto v2 (Ref. BDC-AB/SP01 -> theme B, part of theme C).
  </description_generale>
  <schema_application id="BDCA_CTA"></schema_application>
  ...
  <schema.type_objet nom="ROUTES98_TRON_ROU" type="line">
    <definition>
      Element of the road network. It comprises all the roads, paths and ways of BDCarto. Roads under
      construction are captured only if the relevant bridges and crossroads are already built.
    </definition>
    <schema.type_attribut_ref nomref="#ROUTES98_TRON_ROU_BDC_ID" />
    ...
    <schema.type_attribut_ref nomref="#ROUTES98_TRON_ROU_NBV_TOT" />
    ...
  </schema.type_objet>
  <schema.type_attribut nom="ROUTES98_TRON_ROU_BDC_ID"
    type="entier_naturel" taille="10">
    <definition>Object identifier of BDCarto</definition>
  </schema.type_attribut>
  ...
  <schema.type_attribut nom="ROUTES98_TRON_ROU_NBV_TOT"
    type="enumeration" taille="1">
    <definition>Total number of lanes.</definition>
    <type_valeur_enum code="0">
      Unknown: This value is allowed only for roads under construction or located in urban
      area.</type_valeur_enum>
    <type_valeur_enum code="1">1 lane</type_valeur_enum>
    <type_valeur_enum code="2">2 lanes</type_valeur_enum>
    <type_valeur_enum code="3">3 lanes</type_valeur_enum>
    <type_valeur_enum code="4">4 lanes</type_valeur_enum>
    <type_valeur_enum code="7">2 large lanes</type_valeur_enum>
    <type_valeur_enum code="9">More than 4 lanes</type_valeur_enum>
    <type_valeur_enum code="S">
      Not relevant. This value must be set for road sections with two ways or included in complex
      crossroads.</type_valeur_enum>
    <type_valeur_enum code="A">
      Waiting for source information</type_valeur_enum>
    <type_valeur_enum code="X">Waiting for update</type_valeur_enum>
  </schema.type_attribut>
  ...
  <schema.type_relation nom="RCO_ROUTES98_TRON_ROU"
    type="composition">

```

```

<definition>
Node of the road Network "is located by" node.
</definition>
<schema.type.initial nomref="#ROUTES98_TRON_ROU" />
<schema.type.final nomref="#ROUTES98_ARC" />
</schema.type_relation>
<schema.type_primitive nom="ROUTES98_SOM" type="node">
</schema.type_primitive>
<schema.type_primitive nom="ROUTES98_ARC" type="edge">
</schema.type_primitive>
</Metadonnees.Lot.Definition>
</Metadonnees.Description>

```

Figure 5 Implementation Sample of the French profile CEN/TC 287 prestandard for metadata (dep22bdc20_cta_1998.xml).

As mentioned in section 3, the structure of this XML message relies on the encoding of the European prestandard ENV 12657 (CEN/TC 287, 1999). A specific DTD (XP-ENV12160.dtd) has thus been implemented in order to deliver metadata accordingly to the French profile available in this prestandard. Any other international profile could be easily implemented by defining a specific DTD.

This XML message allows for the delivery of:

- Mandatory (as defined in the CEN/TC 287 prestandard) information about the data set. It includes the name of the theme (ROUTES98), the name of the producer (IGN), the spatial schema type (topology SpatialG1, i.e. Network topology), the language (by default, French), the direct positioning system (LAMBERT), the capture and the last update dates, the co-ordinate unit (cm for centimetre), the indirect positioning system (defined in this example by a rectangular extent).
- The schema of the geographic database, exemplified in Figure 2 by the TRON_ROU class (road section). The class is first defined: Its name, its geometrical feature type (line) and a definition are provided. Each attribute of the class is then detailed (e.g. BDC_ID and NBV_TOT). Relationships are finally defined. In Figure 2, only a pointer to the geometrical structures (from which topological relationships are managed) is presented: the geometry of an TRON_ROU object is defined by sets of nodes and edges.

When the XML files have been processed and the target system is provided with the different metadata, “evolution” messages are delivered. They detail the modifications which have to be performed in the older version of the geographic database. Figure 6 provides an example of such message detailing an evolution undergone by a geographical entity of the road network theme. It deals with a composite (i.e. semantic and geometrical) modification of a road section object (instance of the class TRON_ROU).

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- dep22evol2.xml -->
<!DOCTYPE EvolElem SYSTEM "Evolutions.dtd">
<EvolElem ID="_2">
<Evol.ModifMixte>
<Evol.Evolution TYPE="Identify" OPER="Set">
<Evol.Sem>
<Evol.Reflnst xlink:title="Composite modification: Old Object.">
<Evol.link xlink:href="dep22rte.xml|ROUTES94_TRON_ROU_990063500" />
</Evol.Reflnst>
</Evol.Sem>

```

```

<Evol.Sem>
  <Evol.Reflnst xlink:title="Composite modification: Attribute.">
    <Evol.link xlink:href="dep22rte.xml|ROUTES98_TRON_ROU_990063500_NBV_TOT" />
  </Evol.Reflnst>
</Evol.Sem>
<Evol.Geom xlink:title="Geometrical modification: Primitive.">
  <Evol.link xlink:href="dep22rte.xml|ROUTES98_TRON_ROU_990063500_ARC2" />
</Evol.Geom>
</Evol.Evolution>
</Evol.ModifMixte>
</EvolElem>
<!-- End of evolution dep22evol2.xml -->

```

Figure 6 An example of evolution message (dep22evol2.xml).

The older version of the object involved in the updating operation is first identified (value "Identify" of attribute TYPE of Evol.Evolution). References to the semantic structures (classes and/or attributes) to be modified are then provided (value "Set" of attribute TYPE of Evol.Evolution). In this example, both geometry and semantics are updated. There is only an attribute modification: it concerns the NBV_TOT attribute (total number of lanes). The new value of this attribute can be found in the "geographical application" message (the dep22ret.xml file) by the way of the identifier: ROUTES98_TRON_ROU_990063500_NBV_TOT.

Also a reference to the geometrical primitive of the object involved in the composite modification is provided. As before, the new geometry is to be found in the dep22rte.xml file and identified by ROUTES98_TRON_ROU_990063500_ARC2.

As the implementation of the MD5 algorithm in the prototype is in progress (validation of object components involved in the identifier computation is not achieved and requires peculiar attention in order to uniquely identify the objects), another system of identifiers has been set up for the test. It is based on the existing system of identifiers implemented in BDCARTO®.

Once information about metadata and the nature of the evolution undergone by the objects has been processed by the users' systems, new values are delivered in the "geographical application" message. Figure 7 provides a sample of such a file. This sample focuses on the objects involved in the composite modification previously presented.

```

<?xml version="1.0" encoding="ISO-8859-1" ?>
<!-- dep22rte.xml -->
<!DOCTYPE ApplInfoGeo SYSTEM "ApplGeos.dtd">
<ApplInfoGeo>
...
<!-- Evolution #2 -->
<AIG.Objet ID="ROUTES94_TRON_ROU_990063500"
xlink:href="dep22bdc20_cta_1994.xml#ROUTES94_TRON_ROU">
  <AIG.VAttRef IDREF="|ROUTES94_TRON_ROU_990063500_NBV_TOT" />
</AIG.Objet>
<AIG.VAtt ID="ROUTES98_TRON_ROU_990063500_NBV_TOT"
xlink:href="dep22bdc20_cta_1998.xml#ROUTES98_TRON_ROU_NBV_TOT" VAL="1" />
<AIG.Lien ID=" 2" xlink:href="dep22bdc20_cta_1998.xml#RCO_ROUTES98_TRON_ROU">
  <AIG.ObjRef IDREF="|ROUTES94_TRON_ROU_990063500" />
  <AIG.GeomRef IDREF="|ROUTES98_TRON_ROU_990063500_ARC2" />
</AIG.Lien>
<AIG.Geom ID="ROUTES98_TRON_ROU_990063500_ARC2"
xlink:href="dep22bdc20_cta_1998.xml#ROUTES98_ARC">
  <AIG.Li>
    <AIG.Pt X="22269300" Y="240228900" />

```

```

<AIG.Pt X="22266103" Y="240228100" />
<AIG.Pt X="22233335" Y="240236415" />
<AIG.Pt X="22225611" Y="240238375" />
</AIG.Li>
</AIG.Geom>
...
</AppInfoGeo>
<!-- End of geographical application dep22rte.xml -->

```

Figure 7 Sample of a “geographical application” message (dep22rte.xml).

In accordance with the schemas of the different versions of the geographic database, which have been previously delivered in the metadata messages and with the references provided in the evolution file, the new value of the attribute NBV_TOT of the TRON_ROU instance is delivered (VAL="1"). The new geometry of the object involved in the composite modification is also provided in this file. Each point (AIG.Pt tag) defining the new geometry of the updated object is given.

With this final message, the transfer of updating information is completed and the users' systems are provided with minimal, structured and detailed information concerning the evolutions undergone by the geographical entities. Users' systems are then able to perform the modifications in the databases. Messages confirming or not the integration of the different updates should be sent back to the server, for the consistent synchronisation of the databases (in terms of updating) on both server and client sides.

5 CONCLUSION AND OUTLOOKS

A new delivery mode of updating information between geographical information systems has been presented and detailed. It is designed to provide users with minimal, structured and detailed information on evolutions in order to make their integration in systems easier and more consistent. Based on XML, this solution allows the concerns raised by the lack of interoperability between GIS software to be tackled and to address the problem of real time updating for geographic databases in client/server architectures. Besides, it defines an effective data exchange format for geographic databases: a complete data set is delivered when transferred updates correspond only to creation orders.

The structure proposed in this paper for the delivery of updating information acts as an event based protocol (online or offline) between the server and the clients. Each event corresponds to a type of evolution undergone by geographical objects. It could be accessed through a standard event based API (Application Programming Interface): such an API could be SAX (Simple API for XML). Moreover, the use of such an event based API fulfils the scalability requirements. A non event based API, such as DOM (Document Object Modelling), though easier to implement, would imply the entire XML transfer to be downloaded in memory.

A full-scale implementation of this solution taking advantage of the Java technology and of the web-based client/server data access architectures, as presented in this paper, will be achieved in the near future at IGN in order both to support the real time maintenance of all the databases produced at IGN and to supply consumers with adapted updates on demand.

6 REFERENCES

- Badard, T., 1998. Towards a generic updating tool for geographic databases. In *Proceedings of GIS/LIS'98, Annual Conference and Exposition*, Fort Worth, Texas, USA, pp. 352–363.
- Badard, T., 1999. On the automatic retrieval of updates in geographic databases based on geographic data matching tools. In *Proceedings of the 19th International Cartographic Conference (Ottawa'99)*, ICA/ACI (Eds.), Ottawa, Ontario, Canada, pp. 47–56.
- Bouillé, F., 1987. A survey on the HBDS methodology applied to cartography and land planning. In *Proceedings of EuroCarto VI*, Brno, pp. 20–28.
- CEN/TC 287, 1997. ENV 12160 “Geographic information – Data description – Spatial Schema”, CEN/TC 287, 1997.
- CEN/TC 287, 1999. ENV 12657 “Geographic information – Data description – Metadata”, Rev 165, CEN/TC 287, April 1999.
- International Hydrographic Organisation, 1996. IHO transfer standard for digital hydrographic data - Publication S-57, Edition 3.0, November 1996, 126 pages.
- ISO/TC 211, 1999. CD 15046-18 Geographic information – Part 18: Encoding. ISO/TC 211 N 709 Committee draft for comments, edited by ISO/TC 211 Secretariat, (Oslo: Norwegian Technology Standard Institution), 76 pages.
- Network Working Group, 1992. The MD5 message-digest algorithm. RFC 1321, April 1992, edited by Rivest, R., (Cambridge: MIT). URL: <http://www.cis.ohio-state.edu/htbin/rfc/rfc1321.html> or <http://www.faqs.org/rfcs/rfc1321.html>.
- OpenGIS Consortium, 1999a. Web map server interface specification. OpenGIS project document 99-077r1, edited by Doyle, A., (Needham: International Interfaces inc.), 56 pages.
- OpenGIS Consortium, 1999b. Geography Markup Language (GML). OGC RFC 11 13-Dec-1999, edited by Lake, R., (Vancouver: Galdos Systems Inc.), 37 pages.
- Poupart-Lavoie, G., 1997. Développement d'une méthode de transfert des mises à jour de données à référence spatiale. Mémoire de maîtrise (M. Sc.), Université Laval, Centre de Recherche en Géomatique, Québec, Canada, décembre 1997, 128 pages.
- Raynal, L., 1996. Some elements for modelling updates in topographic databases. In *Proceedings of GIS/LIS'96, Annual Conference and Exposition*, Denver, Colorado, USA, pp. 1223–1232.
- Raynal, L., David, B. and Schorter, G., 1995. Building an OOGIS prototype: experiments with GeO2. In *Proceedings of ACSM/ASPRS Annual Convention and Exposition (AutoCarto12)*, Charlotte, North Carolina, USA, Volume 4, pp. 137–146.
- World Wide Web Consortium, 1998. Extensible Markup Language (XML) 1.0. W3C Recommendation 10-February-1998. URL: <http://www.w3.org/TR/1998/REC-xml-19980210>.

3.3 Compléments

Afin d'apporter au lecteur tous les éléments nécessaires à la compréhension des articles précédents, un certain nombre d'explications absentes dans ceux-ci, ainsi que des compléments d'information sur les sujets qui y sont abordés sont fournis dans cette section. Ces compléments concernent les méthodes d'appariement de données géographiques qui sont impliquées dans le mécanisme d'extraction des mises à jour ainsi que son fonctionnement.

3.3.1 Retour sur l'appariement de données géographiques

Cette section s'attache tout d'abord à justifier le choix de l'utilisation de méthodes d'appariement dans le mécanisme de détection et d'extraction des mises à jour. Ensuite, au travers de la description détaillée du fonctionnement du processus générique d'appariement développé au laboratoire COGIT, et sur lequel se base le mécanisme que nous avons défini et développé, nous justifierons la nécessité de l'établissement de séquences d'appariement dédiées à des applications et aborderons la notion de qualité de lien d'appariement.

3.3.1.1 Justification du choix de méthodes d'appariement

En l'absence de systèmes d'identification fiables et stables (existence d'un identifiant ou d'une clef primaire sur chacun des objets) ou de gestion de l'historique, la seule information permettant d'isoler les changements intervenus dans les bases de données géographiques, est la géométrie des objets. En effet, la localisation, la forme, les relations topologiques et éventuellement les attributs d'un objet peuvent permettre de le discriminer. C'est sur ce principe que repose l'appariement de données géographiques et c'est pourquoi ces méthodes ont été utilisées au sein du mécanisme

3.3.1.2 Fonctionnement détaillé du processus générique d'appariement

D'après la définition donnée dans [Badard, 1998c] (article fourni précédemment, cf. section 3.2.1), l'appariement de données géographiques revient à faire réaliser par l'ordinateur, de la manière la plus automatique possible, l'ensemble des processus complexes que notre cerveau effectue lorsque nous observons deux cartes de la même zone mettant en œuvre des représentations différentes (échelle, symbolisation, etc.) et que nous cherchons à retrouver des éléments se correspondant.

Ce processus est utilisé dans de nombreuses applications manipulant l'information géographique, autres que la mise à jour de bases de données :

- Le regroupement de bases de données juxtaposées [Laurini, 1996],
- Le recalage de données sur un référentiel [Lupien et Moreland, 1987 ; Lynch et Salford, 1985],
- L'intégration de bases de données [MRN, 1992],
- Le contrôle qualité [Brooker, 1995 ; Harvey et Vauglin, 1996 ; Vauglin et Bel Hadj Ali, 1998],

- La superposition de couches d'information pour fusionner les géométries [Demirkesen et Schaffrin, 1996].
- La mise en place de bases de données multi-échelles [Devogele, 1997].

Il s'agit néanmoins d'un processus complexe, nécessitant de nombreuses rétroactions entre les différentes phases le composant. [Devogele, 1997, p. 118] a défini la structure du processus générique d'appariement développé au laboratoire COGIT. Elle peut se décomposer comme illustré dans la Figure 3.1.

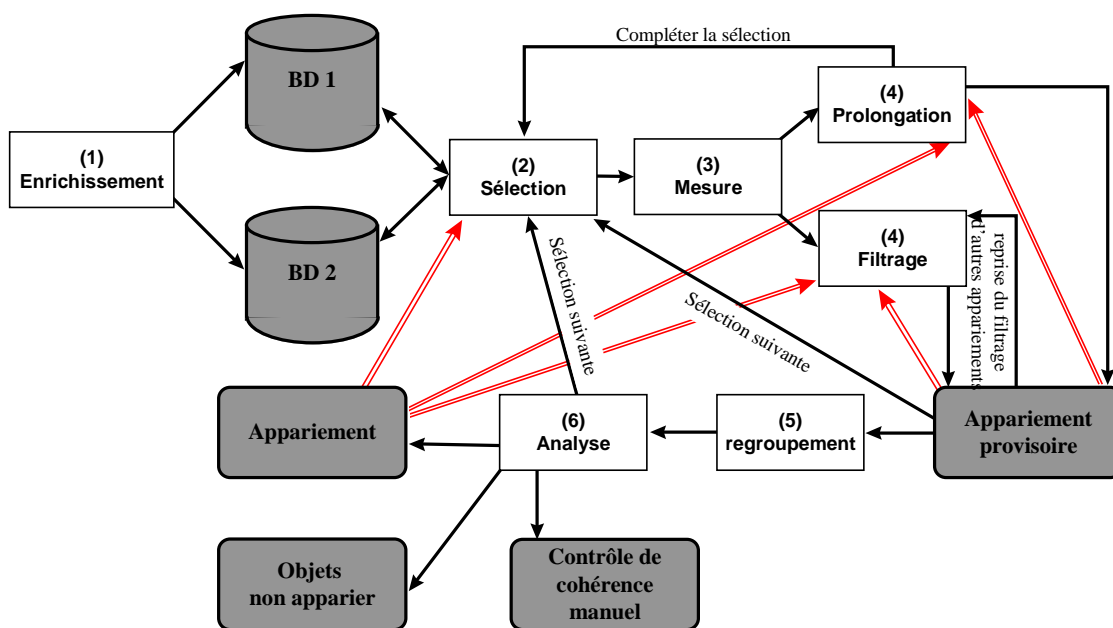


Figure 3.1 - Enchaînement des phases du processus générique d'appariement

Dans un premier temps, les bases de données géographiques à mettre en correspondance sont enrichies par ajout, par exemple, d'attributs virtuels, de méthodes, ou de caractéristiques sur la forme des objets. Cette étape (1) permet d'acquérir des éléments nécessaires au fonctionnement des outils d'appariement (comme les arcs sortants et entrants des nœuds, les faces gauche et droite d'un arc, etc.).

La deuxième étape de ce processus est la sélection (2) : elle désigne dans les deux bases, les groupes d'objets candidats à l'appariement. La sélection des objets candidats permet de restreindre la recherche à une population ciblée afin de limiter le nombre d'éléments parasites et les temps de calcul. Ces sélections peuvent être des populations de classes entières, deux ensembles d'objets répondant aux mêmes critères (ce critère peut être un appariement déjà réalisé), une instance d'une classe de la première base et les instances candidates à l'appariement de la deuxième base en fonction de propriétés de la première instance. L'étape de sélection s'appuie donc sur des outils de sélection qui sont des outils d'appariement rudimentaires (rectangle ou périmètre englobants, distance euclidienne, etc.), sur les appariements déjà réalisés et sur la connaissance des schémas et spécifications des bases de données [Cobb et al., 1998]. En effet, réaliser un appariement sans contraintes sur les classes à appariées est relativement coûteux en temps de calcul et n'est pas de surcroît toujours nécessaire. Il permet cependant de repérer les défauts de classification d'objets qui peuvent intervenir dans les bases de données géographiques au moment de leur saisie.

Un ordre sur les sélections doit être, de plus, établi pour enchaîner les appariements. Les appariements les plus fiables et qui interviennent dans l'établissement d'autres correspondances doivent être exécutés les premiers. Cet ordre va conditionner non seulement la qualité du processus global mais va permettre aussi à un élément candidat à une sélection, qui a été rejeté du processus d'appariement au préalable, d'être sélectionné à nouveau.

Après avoir effectué une sélection, des mesures d'appariement entre les objets sélectionnés ou les objets en relation avec ces derniers sont calculées. La difficulté de cette étape est le choix des outils à utiliser et de leurs paramètres. Tout d'abord, l'utilisateur doit rechercher s'il existe des identifiants communs (ensemble d'attributs ou clef primaire) aux deux bases. Dans l'affirmative, il est souhaitable d'utiliser un outil d'égalité sémantique, le recours à des outils géométriques et topologiques, impliquant des temps de calcul relativement conséquents et qui dépendent de façon non linéaire de la taille des jeux de données à appairer.

Les étapes de filtrage et de prolongation permettent quant à elles, d'interpréter les résultats de la phase précédente et d'affiner les ensembles d'objets candidats à l'appariement.

Le filtrage (3) consiste à retirer les objets parasites des ensembles d'éléments candidats à l'appariement. Ces objets sont détectés soit grâce aux mesures réalisées lors de la phase précédente, soit à l'aide de nouveaux outils (exemple : plus court ou plus proche chemin dans le cas d'appariement de réseaux). Ainsi, les objets ne répondant pas aux critères fixés par l'utilisateur sont supprimés. Le choix des valeurs de ces critères est réalisé, soit en ayant recours aux métadonnées, si elles existent – et notamment aux valeurs des indicateurs de qualité sur la position et la forme des objets – soit empiriquement par des essais successifs. Le filtrage peut aussi être utilisé pour affiner des appariements déjà réalisés. En effet, un processus d'appariement (appelé appariement provisoire) peut être repris une fois qu'un autre processus aura été effectué. Cette consolidation d'appariement provisoire par d'autres appariements est très utile pour rendre fiable l'appariement des objets en relation.

La prolongation (4) consiste à compléter une sélection si celle-ci, au vu des mesures d'appariement, est jugée incomplète pour en déduire un appariement fiable. Pour ces nouveaux objets sélectionnés, la phase de mesure d'appariement doit être lancée. Par exemple, pour appairer les tronçons de réseaux, si les mesures réalisées indiquent que la sélection est incomplète, le plus petit des deux chemins est prolongé en ajoutant une des arêtes connexes, afin d'établir, entre les deux bases, des liens n - m (i.e., n objets correspondent à m objets).

La phase de regroupement des appariements (5) permet de regrouper les résultats obtenus pour retrouver les appariements entre objets. Cette phase est simple et consiste à confronter les différents regroupements obtenus, afin de détecter les objets apparaissant dans plusieurs groupes. Elle est réalisée une fois que tous les objets susceptibles d'être regroupés sont appariés provisoirement.

Enfin, les étapes d'analyse du résultat (6) et de contrôle permettent de s'assurer que toutes les contraintes de cohérence ont été vérifiées. Ainsi, avant de valider les résultats obtenus, il faut analyser les correspondances afin de vérifier la validité des contraintes non encore employées. Ces contraintes peuvent porter, par exemple, sur la cardinalité des appariements obtenus, la connexité, la complétude, le graphe de communication, les relations de composition, etc. Si, les objets vérifient les contraintes déterminées, l'appariement pourra être considéré comme valide. Par contre, si les objets en

correspondance ne les vérifient pas, un contrôle de cohérence manuel, mais guidé par le processus, devra être réalisé.

La description de ce processus générique d'appariement fait tout d'abord apparaître la nécessité de la définition de séquences et d'algorithmes spécifiques en fonction de l'utilisation visée des résultats qu'il produit. Un autre point qu'il met en exergue est lié à la nature même de ces résultats, à savoir des entités liées ou non. Aucune indication sur la fiabilité de ceux-ci n'est fournie. Ces différents points sont abordés dans les sections suivantes.

3.3.1.2.1 Nécessité d'un appariement adapté à l'extraction des mises à jour

La description du processus générique d'appariement, réalisée précédemment, fait en effet apparaître que non seulement l'ordre des différentes sélections opérées mais aussi le choix des algorithmes sur lesquels il repose, dépendent fortement du but visé par cette mise en correspondance de jeux de données, c'est-à-dire de la motivation qui a conduit à réaliser l'opération d'appariement. Or, l'ajout d'éléments, basés par exemple sur des systèmes à base de règles et de contraintes, tels qu'utilisés dans le processus de généralisation proposé par [Ruas, 1999], de multi-agents [Ferber, 1995 ; Demazeau, 1995] ou d'apprentissage [Michalski et al., 1984], qui permettent de capitaliser et reproduire la connaissance des experts, afin de doter le processus d'un comportement intelligent, adaptatif et autonome, le rendant ainsi capable de réaliser l'appariement en fonction de son utilisation visée, sont toujours du domaine de la recherche (notamment au laboratoire COGIT). Ceci nous a donc obligé à développer une séquence propre et des algorithmes d'appariement spécifiques, afin de permettre la mise en correspondance de deux actualités d'une même base de données géographiques, le but étant d'en extraire les mises à jour.

Néanmoins, peu de travaux, dans la littérature, se sont attachés à définir des méthodes d'appariement adaptées à l'extraction des évolutions dans les bases de données vecteurs qui soient réellement génériques, c'est-à-dire, capables de traiter tous les types sémantiques et géométriques d'entités géographiques et ne nécessitant pas l'existence d'un système d'identification. Ces travaux ont, en effet, principalement porté sur le développement d'outils spécifiques permettant la recherche de cette information pour certains types de données uniquement. Ainsi, [Gis/Trans Ltd, 1995] a défini un certain nombre d'outils principalement basés sur la sémantique, [Bucaille, 1996] a développé un ensemble d'algorithmes géométriques et topologiques afin de traiter les données routières et [van Oosterom, 1994] a conçu une méthode basée sur la topologie qui permet de retrouver les éléments correspondants entre deux couches de bâti présentant des actualités différentes.

Bien que spécifiques, ces travaux ont mis en évidence un point important permettant de rendre le processus d'appariement dédié à l'extraction des évolutions, réellement générique. Ils démontrent, en effet, qu'il est nécessaire de s'appuyer sur la géométrie et la topologie des données afin de s'absoudre de leur type sémantique et de pouvoir détecter les défauts de classification des objets, les bases de données géographiques n'étant malheureusement pas exemptes de ce type d'erreur. Cela impose donc de disposer d'une bibliothèque d'algorithmes dédiés permettant de traiter tous les types de primitives géométriques et de retrouver leurs correspondances, avant de pouvoir établir les appariements entre objets (sémantiques). C'est sur ce principe que nous avons basé le processus d'appariement mis au point et développé au cours de ce travail de thèse, et qui est décrit dans [Badard, 1998c] (article fourni en section 3.2).

3.3.1.2.2 Notion de qualité des liens d'appariement

Le processus d'appariement de données géographiques, comme le montre la description détaillée de son fonctionnement, fournit actuellement des résultats de deux types : des objets ou des groupes d'objets appariés (i.e., en correspondance), et des objets isolés (i.e., sans homologues dans l'une des deux bases mises en jeu). Même si ce processus devient de plus en plus robuste, rien ne vient donc qualifier « la fiabilité » avec laquelle chacune de ces correspondances a été établie. On entend par là, le niveau de confiance que l'on peut accorder à l'établissement ou non d'un lien d'appariement. Il s'agit pourtant d'une information cruciale, dès que l'on réalise des applications s'appuyant sur ce genre processus (bases de données multi-échelles, fusion de géométrie, contrôle qualité, et bien entendu mise à jour). En effet, toute action menée dans de tels systèmes et utilisant une information de correspondance erronée peut avoir des conséquences irrémédiables sur l'état de cohérence des informations qu'ils stockent ou sur la véracité des résultats obtenus.

Bien loin d'avoir apporté des éléments de réponse, s'appuyant sur un cadre formel et théorique tels que l'utilisation de logiques de défaut, voire d'éléments de logique flou [Gacogne, 1997], nous avons défini dans le cadre de nos recherches un certain nombre d'indicateurs qui permettent de fournir une première estimation de cette notion de qualité des correspondances établies. Ces indicateurs rendent, en effet, possible la détection des objets au sein d'un groupe dont l'appariement paraît suspect. Ils sont basés principalement sur des mesures de position relative (par exemple, la mesure de la distance euclidienne entre les différents constituants), d'écarts de forme (distance de Hausdorff, distance de Fréchet, etc.) et de différence de niveau de représentation des objets ou des groupes d'objets intervenants dans un appariement (indicateur de contraste sur le nombre d'objets impliqués, taux de fragmentation géométrique, etc.). Confrontés aux valeurs des indicateurs de qualité sur la position et la forme des objets fournies en général dans les métadonnées des jeux mis en correspondance, ces différents indicateurs permettent la détection de liens potentiellement erronés.

L'intégration de ces mesures de qualité au sein même du processus peut donc améliorer la qualité des résultats qu'il produit. Néanmoins, elle pose le problème de la gestion de la propagation et de l'estimation de cette information au cours du processus. Les travaux actuellement engagés par [Bonin, 1998] sur l'étude de la propagation des indicateurs de qualité de la donnée géographique au travers de certaines applications, dans le but d'estimer des indicateurs pertinents vis à vis de celle-ci, apporteront certainement des éléments de réponse à ce problème.

3.3.2 Retour sur le mécanisme d'extraction des mises à jour

Deux compléments relatifs au processus d'extraction des mises à jour sont ici détaillés. Ils concernent les problèmes de robustesse du processus d'appariement sur lequel il s'appuie et la notion de « traçabilité » de l'information d'évolution.

3.3.2.1 Problèmes de robustesse du processus d'appariement

La notion de qualité des liens d'appariement décrite dans la section 3.3.1.2.2 renvoie, pour partie, à la façon dont le processus d'appariement fonctionne. Les méthodes de recherche des objets homologues et de validation des correspondances établies sont, en effet, basées sur des seuils de tolérance. Elles ne passent donc pas par

la définition et la formulation d'une fonction mathématique pour laquelle le but à atteindre (i.e. un appariement valide) consiste en la recherche d'un optimum (exemple : extraction de sous graphes de cardinalité minimale). Il en résulte ainsi l'apparition de liens parasites au cours du processus, voire même d'appariements parfois incomplets.

Ce manque relatif de robustesse du processus d'appariement ne vient cependant pas entacher la complétude du mécanisme d'extraction des mises à jour. En effet, ce processus traite les données dans leur intégralité. Ainsi, même si un lien de correspondance incomplet est établi lors de la phase d'appariement, tout changement sera tout de même détecté, seule la nature de l'évolution ne sera pas exacte. Ce processus est donc complet, c'est-à-dire qu'il ne laisse échapper aucune modification des jeux de données, seule l'appartenance à un type d'évolution varie en fonction des seuils retenus lors des différentes étapes d'appariement. Ce processus nécessite donc un choix judicieux des différents seuils impliqués, si l'on désire accéder à une classification des mises à jour proche des évolutions réellement subies par les entités géographiques.

La mise en œuvre de méthodes d'optimisation, par exemple tirées de la théorie des graphes, serait néanmoins à investiguer pour améliorer la robustesse du processus d'appariement et ainsi diminuer les cas de liens de correspondances erronés ou incomplets.

3.3.2.2 Notion de « traçabilité » de l'information d'évolution

Etroitement lié aux problèmes de classification de la nature des évolutions fournies par le mécanisme d'extraction des mises à jour, se pose celui de la « traçabilité » de cette information. En effet, si l'on accède bien à une description détaillée des mises à jour intervenues entre deux actualités, on ne dispose pas des raisons ayant entraîné ces changements. S'il semble difficilement envisageable de pouvoir retrouver cette information par des méthodes de comparaison de jeux de données telles que celle développée, il s'agit néanmoins d'une information importante qu'il serait utile de mettre à disposition afin de mieux appréhender la pertinence des mises à jour opérées sur le référentiel et de mieux en contrôler les effets dans les systèmes.

Le producteur de données, réalisant les mises à jour de les bases de données de référence et disposant donc de cette information, pourrait la fournir dans ces jeux de données ou dans ces lots d'information d'évolution, via le renseignement de métadonnées spécifiques. La définition et l'utilisation de telles métadonnées a déjà été investiguée dans le cadre du Plan Cadastral Informatisé par [Spéry, 1999].

Néanmoins, il semble qu'il y ait autant de « traçabilités » différentes qu'il y a de types d'utilisation de la donnée géographique. On se heurte donc aux problèmes d'interopérabilité sémantique de l'information géographique [Kuhn, 1994 ; Bishr, 1998]. Si la mise en œuvre de ces métadonnées dédiées est un jour entreprise, elle devra passer par une démarche de normalisation préalable afin de pouvoir être utilisée par tous les utilisateurs.

3.4 Résultats des cas d'étude

Cette section présente les résultats de l'extraction des mises à jour sur les deux cas d'études qui ont été présentés dans la section 2.6.

Afin de retrouver les évolutions subies par la BDCarto et la BDTopo entre les deux actualités dont nous disposons, nous avons appliqué le mécanisme d'extraction des mises à jour défini dans ce chapitre. Les résultats sont fournis et illustrés dans les deux sections suivantes.

3.4.1 Extraction des mises à jour de la BDCarto

La Figure 3.2, fournie ci-dessous, illustre les résultats obtenus par le module d'extraction des mises à jour. Dans toute la suite de ce document, on appellera BDCARTOV1, le jeu de données BDCarto de 1994 et BDCARTOV2, la version de la BDCarto datant de 1998. De même, le jeu Route500 ancien sera nommé R500V1. Le jeu Route500, à jour et produit par l'IGN, s'appellera alors R500V2.

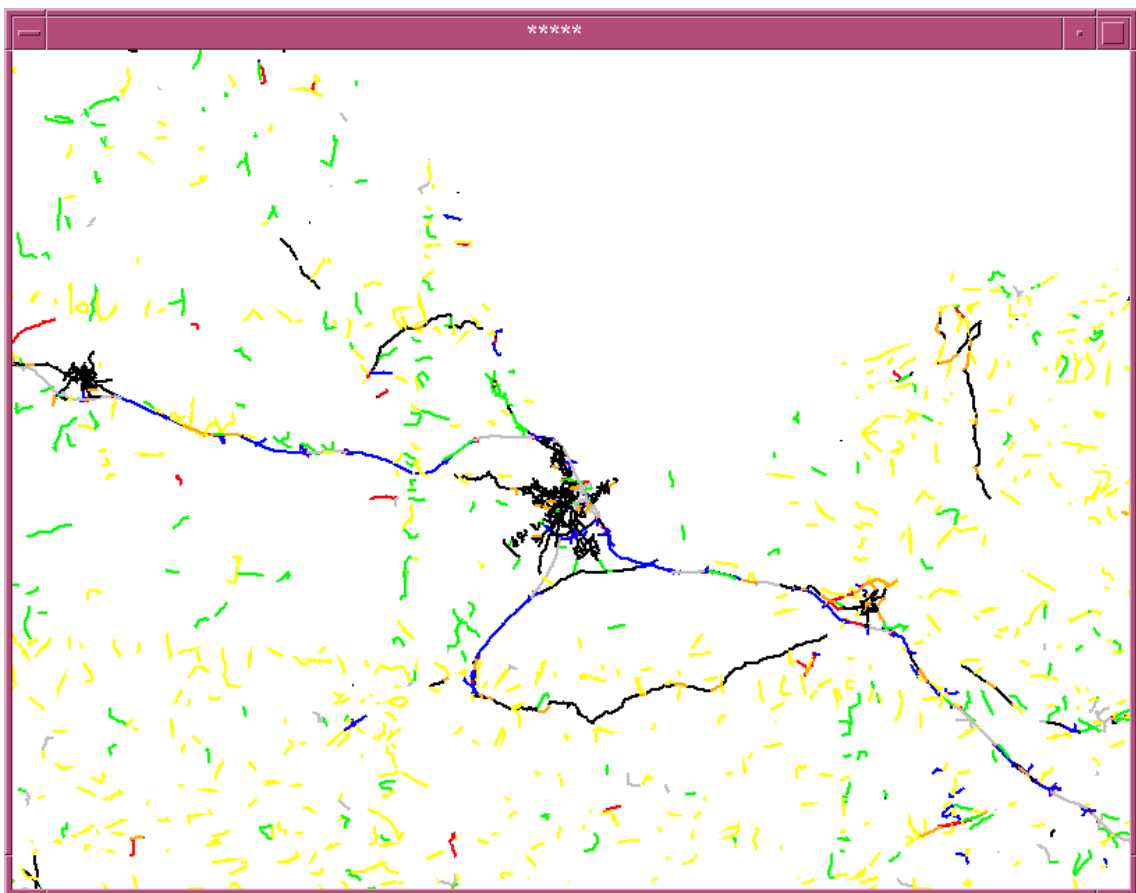


Figure 3.2 - Calque différentiel (objets contenus dans BDCARTOV2)

Légende :

- ROUGE : les objets créés.
- BLEU : les objets provenant d'une scission.
- VERT : les objets provenant d'une fusion.
- GRIS : les objets provenant d'une agrégation.
- NOIR : les objets ayant subi un changement sémantique.
- JAUNE : les objets ayant subi un changement de géométrie.
- CYAN : les objets ayant subi un changement de classe
- MAGENTA : les objets ayant subi un changement de classe ainsi qu'un changement de géométrie

Les résultats détaillés de l'extraction sont donnés ci-dessous

Type de mise à jour	Nombre d'opérations de mise à jour	Nombre d'objets ayant évolués dans BDCARTOV1	Nombre d'objets mis à jour dans BDCARTOV2
<i>Création</i>	1557	0	1557
<i>Destruction</i>	1482	1482	0
<i>Changement de classe</i>	0	0	0
<i>Changement de classe et de géométrie</i>	0	0	0
<i>Modification sémantique</i>	1658	1658	1658
<i>Modification sémantique et géométrique</i>	186	186	186
<i>Modification géométrique</i>	2222	2222	2222
<i>Agrégation</i>	142	337	342
<i>Scission</i>	316	316	679
<i>Fusion</i>	904	1941	904
<i>Total des mises à jour</i>	8467	8142	7548
<i>Nombre total d'objets</i>		64303	63709
<i>Taux de mise à jour</i>		13 %	12 %

Tableau 3.1 - Résultats détaillés de l'extraction des mises à jour entre les deux versions de la BDCarto

3.4.2 Extraction des mises à jour de la BDTopo

La Figure 3.3, fournie ci-dessous, illustre les résultats obtenus par le mécanisme d'extraction des mises à jour appliqué à la BDTopo. Dans toute la suite de ce document, on appellera BDTOPOV1, le jeu de données BDTopo de 1994 et BDTOPOV2, la version de la BDTopo datant de 1996. De même, la BDTopo25 ancienne sera nommée TOP25V1. Le jeu BDTopo25, à jour et produit par l'IGN, s'appellera alors TOP25V2.

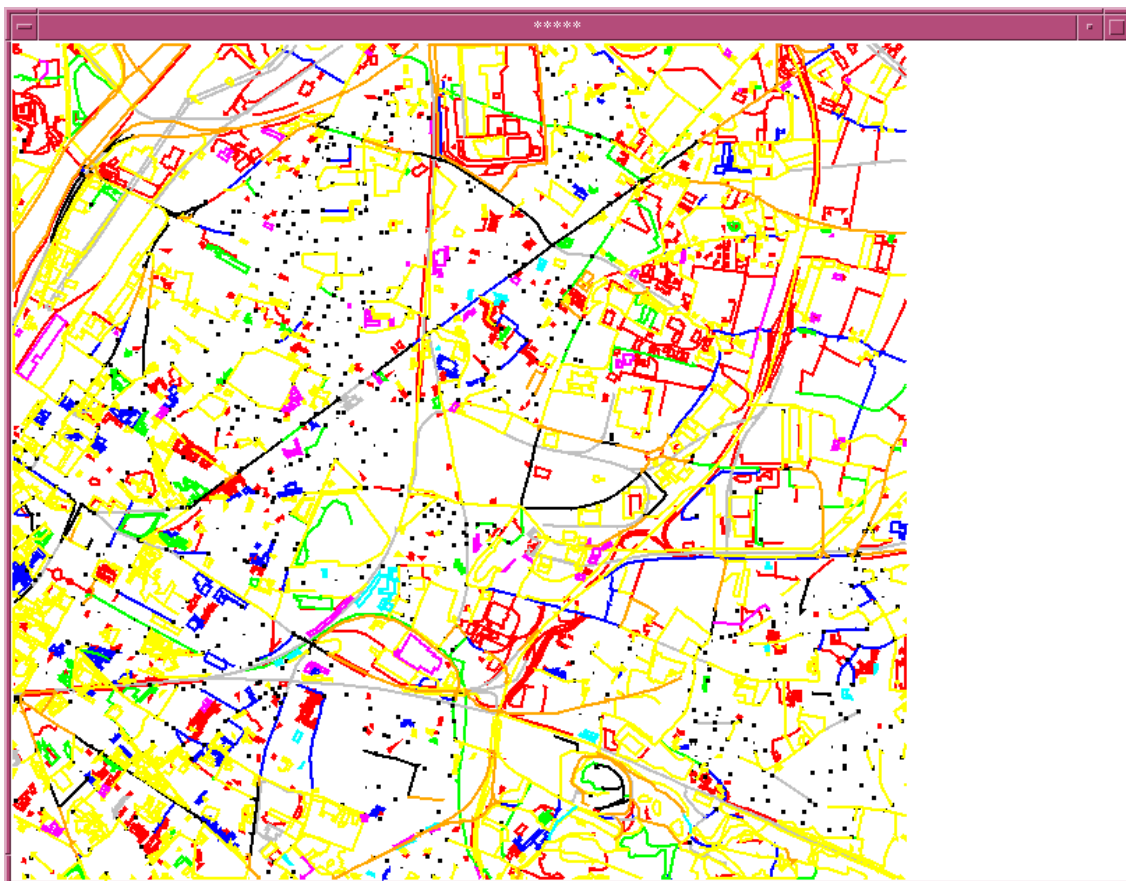


Figure 3.3 – Calque différentiel (objets contenus dans BDTPOV2)

Légende :

- ROUGE : les objets créés.
- BLEU : les objets provenant d'une scission.
- VERT : les objets provenant d'une fusion.
- GRIS : les objets provenant d'une agrégation.
- NOIR : les objets ayant subi un changement sémantique.
- JAUNE : les objets ayant subi un changement de géométrie.
- CYAN : les objets ayant subi un changement de classe.
- MAGENTA : les objets ayant subi un changement de classe ainsi qu'un changement de géométrie.

Les résultats détaillés de l'extraction sont donnés ci-dessous :

Type de mise à jour	Nombre d'opérations de mise à jour	Nombre d'objets ayant évolués dans BDTPOV1	Nombre d'objets mis à jour dans BDTPOV2
<i>Création</i>	1298	0	1298
<i>Destruction</i>	566	566	0
<i>Changement de classe</i>	30	30	30
<i>Changement de classe dû aux modifications de spécification</i>	4	4	4
<i>Changement de classe et de géométrie</i>	61	61	61
<i>Modification sémantique</i>	1222	1222	1222
<i>Modification sémantique due uniquement aux modifications de spécification</i>	1105	1105	1105
<i>Modification sémantique et géométrique</i>	99	99	99
<i>Modification géométrique</i>	1002	1002	1002
<i>Agrégation</i>	50	146	162
<i>Scission</i>	128	130	309
<i>Fusion</i>	92	205	90
<i>Total des mises à jour</i>	4548	3461	4273
<i>Nombre total d'objets</i>		9079	9891
<i>Taux de mise à jour</i>		38 %	43 %
<i>Taux de mise à jour dues aux changements de spécifications</i>		12 %	11 %
<i>Taux de mise à jour dues à l'évolution du terrain</i>		26 %	32 %

Tableau 3.2 –Résultats détaillés de l'extraction des mises à jour entre les deux versions de la BDTopo d'Angers

Chapitre 4

Intégration et propagation des mises à jour

4.1 Introduction

Ce chapitre présente le processus d'intégration et de propagation des mises à jour dans les bases de données géographiques qui est au cœur du mécanisme proposé dans ce mémoire. L'ensemble des détails concernant sa définition et son fonctionnement (notamment les procédés mis en place pour permettre la conservation des acquis et de la cohérence dans les systèmes), est donné en section 4.2. Un certain nombre de compléments, relatifs au modèle de gestion sur lequel repose ce processus et sur la stratégie de gestion des conflits de mise à jour qui peuvent survenir sont fournis dans la section 4.3. Les résultats obtenus sur les cas d'étude sont fournis et discutés dans la section 4.4.

4.2 Contributions

L'article référencé ci-dessous constitue la contribution apportée à la résolution du problème d'intégration et de propagation de l'information d'évolution dans les bases de données géographiques. Cet article est fourni dans les pages qui suivent.

[Badard et Lemarié, 1999] T. Badard et C. Lemarié. Propagating updates between geographic databases with different scales. Chapter 10 of Innovations in GIS VII: GeoComputation, Atkinson, P. and Martin, D. (Eds.), Taylor and Francis, London, 1999, 12 pages.

Propagating updates between geographic databases with different scales

Thierry Badard and Cécile Lemarié

10.1 INTRODUCTION

Nowadays, Geographic Information Systems are considered to be truly analysis and decision-making tools. For that reason, a growing number of organisations invest in such systems and add specific information necessary to the tasks for which they have the responsibility. However, the implementation of such systems is difficult and relatively expensive. That is why these institutions purchase reference data sets from geographic information producers. From these they are able to develop their own information systems. Thus, these organisations are considered, by the producer, as users of reference geographic data sets.

To carry out their assignments, these institutions (e.g. governmental agencies—Department of Public Works, Environmental Agency—, public utilities—electricity, gas, water—, private companies with a national remit, etc.) clearly need updates from the producer, in order to have the most faithful and realistic image of geographic reality. But these updates must allow for the preservation of consistency of all the knowledge added in the systems by users. Producers (e.g. National Mapping Agency) may have similar preoccupations if they want to propagate the updates from their reference data sets to their derived products (e.g. cartographic products or other databases with different scales) in order to reduce the production costs. However, at the present time, such a propagation of the effects of spatio-temporal evolutions from a reference database to a user's or a derived geographic database, may induce significant risks of information loss or leave the database in an inconsistent state. A taxonomy of the problems hindering the integration of the updates for geographic databases has already been presented and detailed in Badard (1998).

This paper focuses on a generic updating tool for geographic databases under development at the COGIT laboratory of IGN (the French National Mapping Agency), which takes all these problems into account. This mechanism aims at automating the propagation of updates between databases with possible different levels of abstraction, representation and scale.

In section 10.2, some interesting experiences dealing with the updating of geographic databases are presented in order to point up the different elements necessary to the definition of a generic updating tool. The structure of such a generic mechanism is proposed and detailed in section 10.3. Section 10.4 focuses on the propagation step of this process and details how the effects of the updates on the different representations are controlled. In order to illustrate this updating process, examples stemming from two experiments are presented and discussed in section 10.5. The first one concerns the propagation of updates from BDTPOPO® (a 1:25,000 topographic database with 1-meter resolution produced by the IGN) to its derived cartographic database. The second one is about the updating of Route500®, a geographic database stemming from a generalisation

of the road network of BDCARTO® (a 1:100,000 database with 10-meter resolution produced by the IGN).

10.2 SOME EXPERIENCES WITH THE UPDATING OF GEOGRAPHIC DATABASES

The purpose of this section is not to draw up a complete review of the different solutions which have been implemented to tackle the crucial issue of updating in multi-scale databases. Most of the approaches presented in the literature fulfil specific needs, hypotheses and constraints which often are not fully compliant with the requirements of a geographic data producer responsible for the maintenance of its different databases. Nevertheless, two interesting methods are detailed here because they allow the identification of the different elements necessary to the definition of a generic mechanism dedicated to the updating of multi-scale databases.

Thus, the mechanism for the propagation of updates in multi-scale databases defined in Kilpeläinen (1995) provides an interesting example of what a generic method could be. This mechanism relies on the assumption that each generalisation process involved in the derivation of the different representations is formally defined. This is what she named an *incremental generalisation environment*. In such an environment, updating is performed as detailed below:

“In an incremental generalisation environment the generalisation process is done completely for the whole geodata base only once. After the following update transactions to the geodata base, the old generalised output is also updated in an incremental way, which means that the generalisation process is performed only for the modules influenced by the updates” (Kilpeläinen, 1995).

This updating mechanism suits the specific case of multi-scale databases where the generalisation processes between all representations are explicitly defined and controlled. Nevertheless, the preservation of consistency in the different representations relies on a relevant identification of the *modules* which have been updated. This term is not precisely defined because it depends on the generalisation processes involved in the propagation. A *module* includes not only the modified objects in the base level but also all contextual information necessary to the production of an up-to-date representation compliant with the whole generalisation process previously performed. Indeed, if the set of selected entities is not complete, the propagation process may result in inconsistencies: effects of the update on the contextual information may not be completely performed. So, the identification of *modules* is a crucial point of the propagation process but it is far from being easy to implement.

Another interesting approach is defined in Uitermark *et al.* (1998). It deals with the construction of a prototype for propagating building updates between two topographic maps with different scales. This prototype is based on an update propagation strategy that can be decomposed in six steps:

1. The large-scale database is rolled back in time till the moment of actuality of the mid-scale data set.
2. Correspondences between the two data sets are determined via a map overlay process.
3. Correspondences are checked for inconsistencies due to the different specifications of the two data sets.

4. Aggregates are reconstructed and updates are filtered with respect to the mid-scale database specifications.
5. The relevant updated entities are generalised.
6. The generalised objects are integrated in the mid-scale database by means of a rubber-sheeting adjustment algorithm.

Although this process is limited to the updating of buildings, it emphasises the relevance of defining a strategy to perform a consistent and complete propagation of updates between databases with different scales. However, as in Kilpeläinen (1995), spatial integrity is not fully controlled (e.g. intersections with another type of geographical entities may occur and are not detected by the process), which may leave the database in an inconsistent state.

Moreover, in both processes the updates are clearly identified, which is far from being a general case. Indeed, only few stable and reliable systems of identifiers or time management are implemented in geographic databases. Therefore, a mechanism for the updating of multi-scale databases to be considered as truly generic must include a process for the automatic retrieval of updates between two versions of the same geographic database.

Considering the different elements pointed up in this section, a generic mechanism dedicated to the propagation of updates in multi-scale databases can be defined. The structure of such a mechanism is detailed in the next section.

10.3 STRUCTURE OF THE UPDATING MECHANISM

As mentioned in the previous section, this generic mechanism allows the updating of geographic databases from the retrieval of changes in the reference data sets to the integration of the evolutions and to the propagation of their effects in derived or users' databases. The general structure of this mechanism is illustrated in Figure 10.1:

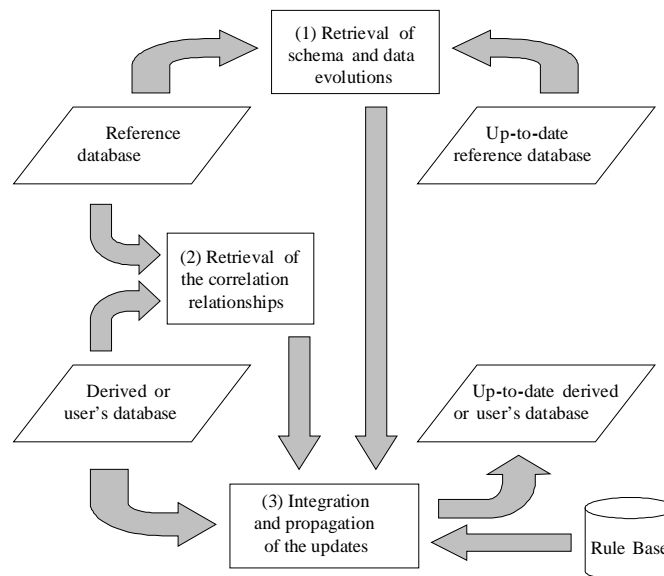


Figure 10.1 Structure of the updating mechanism for geographic databases.

As illustrated in Figure 10.1, this mechanism may be divided in three main steps. The first step consists in the automatic retrieval of schema and data evolutions between two versions of the reference database (Badard, 1999). This step may be divided in two main sub-processes.

A geographic data matching process is first performed. Geographic data matching is akin to the notion of conflation defined in Saalfeld (1988) and Laurini and Thompson (1992). It consists in the computation of correspondence relationships between sets of geographical entities that represent the same phenomenon in two representations of the real world (Lupien and Moreland, 1987; Lemarié and Raynal, 1996; Lemarié and Bucaille, 1998). The two representations may have very different scales, levels of abstraction and/or representation. Such a mechanism has been developed at the COGIT laboratory of IGN (the French national mapping agency) for several years. It relies on the implementation of numerous algorithms involving all geographical information levels (i.e. semantic, geometrical and topologic) and allows the processing of all types of geographical feature: points, lines and areas.

Finally, the correspondence relationships processed between the two versions of the reference database are analysed in order to retrieve updates. This analysis deals with the cardinalities of the relationships and in accordance with the typology of the spatio-temporal evolutions defined in Badard (1999), objects which have been created, deleted, merged, split, aggregated or semantically and/or geometrically modified are detected and extracted.

The second step deals with the preservation of the spatial integrity constraints. It aims at retrieving often implicit but relevant relationships between spatial entities in order to perform a complete and consistent updating of geographic databases. This minimal set of necessary relationships that we have named *correlation relationships* are established not only between the databases but also within the data sets themselves. They can be classified into four main categories as proposed in Badard (1998):

- The composition relationships: An entity is composed of several sub-entities (e.g. a road is comprised of road sections).
- The correspondence relationships: These are established between sets of entities that represent the same phenomenon in two representations of the real world.
- The geometrical relationships: These are meant to translate spatial relationships between geographical entities and are especially concerned with the notions of geometry sharing, intersection, inclusion and connection as defined for instance in Egenhofer and Herring (1990) and Lemon and Masters (1997).
- The dependency relationships: These are relationships as established between objects designed in “classical” databases. They are used to compute attribute values. For geographic databases, the geometry may be considered as an attribute.

This step involves mechanisms partially based on the same generic data matching toolbox used in the first step, which generates a set of geometrical and topologic relationships. Such relations are analysed and classified by means of the semantics of the geographical entities. In order to automate this extraction process, specific metadata describing the nature of entities, the scales, and the thresholds of detection have been defined (Bonnani, 1998).

The third step deals with the updating of users’ or derived databases. The correlation relationships previously extracted are used in order to propagate the effects of the updates detected in the reference data sets. To control the propagation process and allow a complete and consistent updating of geographic databases, a rule base is

implemented. Such a rule base takes the available specifications and metadata (scales, resolution, ...) into account and triggers the appropriate algorithms for the updating of the geographical entities stored in the databases. The rule base is fully upgradable and then allows the management of specific conflicts. Unsolved conflicts are listed by the process and can be solved either interactively or by enhancing the rule base.

This final step is thus the core of the updating mechanism. Very different sources of information (types of update, correlation relationships, metadata, rules of propagation, ...) and algorithms are involved in it. In order to better understand how these different sources of information interact and how algorithms are triggered, the propagation step of the updating mechanism is detailed in the next section.

10.4 THE UPDATING MECHANISM

As mentioned before, this step follows the retrieval of the updates and of the different relationships necessary to a complete and consistent updating of the data sets. Due to the differences between the databases in terms of content and structure, this step can be very difficult to perform. As the different modifications performed by users in their databases can be considered as a derivation process of the reference database, in this section, the term *derived database* will refer either to a derived or a user's database. Most of the difficulties that the propagation process has to take into account, result from:

- Differences between schemas of the databases: two classes in the reference database may have for instance been merged in a single class in the derived or user's database by means of an attribute.
- Differences between data stored in the databases: users may have already integrated some updates in their databases or have added specific information, which are not included in the reference data set.
- Differences between the levels of representation: databases stemming from a derivation process (e.g. generalisation) do not include the same data as in the reference data set.

Despite these difficulties, an updating mechanism which performs a complete and consistent updating of any kind of derived or users' databases has been defined and its implementation is in progress. The process can be decomposed in five main steps:

1. The update filtering;
2. The integration of the updates;
3. The propagation of the effects of updates;
4. The management of updating conflicts;
5. The checking of the database consistency.

These different steps are presented and detailed in the following sections.

10.4.1 Update filtering

This step consists in the selection of the relevant updates which have to be integrated in the derived database. This step is very important because each modification performed in the derived database may induce conflicts with the data modified or added by the user. It is then necessary to narrow the integration of the updates in the derived database only to

the objects which need to be updated. To select which modifications have to be integrated in the derived database, the following information is used:

- the available metadata of the derived database to determine if the modification is relevant in comparison with the resolution of the database. For instance it is not necessary to integrate a geometric modification of 10 centimetres in a derived database where the geometric resolution is estimated to 10 meters.
- the derivation process (when it is formally defined) in order to check if a modification in the reference database has an impact in the derived database. For instance it is not worth integrating a splitting due to the modification of an attribute value if this attribute is not in the derived database.
- the correspondence relationships which have been extracted between the reference database and the derived database to determine if the updated object in the reference database is represented in the derived data set.

In order to automate this filtering step, all selection criteria are expressed by means of rules and stored in a rule base. Such a rule base is fully upgradable and allows not only automation of the selection process but also to control the integration and the propagation of the updates. These different steps are detailed in the following sections.

10.4.2 Update integration

Once the updates have been filtered, each relevant update has to be integrated in the derived database. Each update is processed separately. First, the objects which have to be updated in the derived database are detected by means of the correspondence relationships extracted during the correlation relationships retrieval step. These objects from the derived data set are in correspondence with objects from the reference data set involved in the update. Update integration in the derived database depends on:

- The type of evolution (deletion, creation, splitting, merging, aggregation, semantic—i.e. attribute and/or class—modification, or geometrical modification).
- The cardinality of the correspondence relationships between the reference and the derived data sets (i.e. 1-to-1, 1-to-n, n-to-1 or n-to-m).
- The derivation process between the reference and the derived data sets.

These three types of information determine which algorithms have to be triggered in order to integrate the updates in the derived database. The first and the second one are independent of the derived and the reference data sets whereas the derivation process is completely dependent on this couple of databases. But, two situations may occur:

- The derivation process may be formally defined: then it is possible to entirely express it by means of integration rules and a fully automated integration of the updates can be performed.
- The derivation process may be partially determined because manual interactions are involved: it is then difficult to express the derivation process by means of integration rules in order to automate the whole step of update integration. As a result, a manual interaction may be required in the integration step in order to process the unexpressed part of the derivation process.

Therefore, the rule base previously defined is enhanced by including some generic rules independent of the reference and the derived databases and some specific rules which express the derivation process between both data sets. For the expression of the integration rules, two strategies are possible. The derived objects are updated by applying the same type of update as was detected in the reference database. Otherwise, the derived objects are updated by performing anew the derivation process on the up-to-date objects which are stored in the reference database.

The first solution is implemented when the derivation process is not fully determined (i.e. manual interaction is involved) but partially expressed by means of integration rules. The second solution is preferred when the derivation process is formally defined by the integration rules. The latter solution allows the preservation of the consistency of the derived database. Nevertheless, these two strategies are often combined in order to apply the most appropriate method to each type of evolution.

Most of the time, the expression of the integration rules allows the automation of an important part of the integration process. Specific updates for which it has not been possible to define integration rules, are processed manually as described in section 10.4.4.

10.4.3 Update propagation

Once the updates have been integrated in the derived database, the objects which are correlated with the updated entities are processed. This step which aims at propagating effects of the updates relies on two processes:

1. The retrieval of objects which are correlated with the up-to-date objects within the derived database. This retrieval is performed by using the correlation relationships which have been computed in the second step of the whole process (cf. section 10.3).
2. The update propagation. Two strategies can also be defined depending on the knowledge of the derivation process. If the derivation process is formally defined, then it is possible to determine if the effects of the update have to be propagated to the correlated objects. Rules of propagation are then expressed to define how the propagation is performed. If the derivation process is partially determined, constraints which have to be checked between entities in order to maintain spatial integrity can be expressed. But these constraints do not necessarily imply that the propagation rules are fully defined. So, some correlated objects will not be automatically updated but the process allows for their detection. These unsolved updates are processed interactively by the user as described in section 10.4.4.

10.4.4 The management of updating conflicts

The previous steps have demonstrated that an important part of the process can be automated. Nevertheless, the process has often to deal with specific integration conflicts. Due to the specificity of some updates or the lack of relevant integration and/or propagation rules, the propagation and/or integration processes of updates may result in inconsistencies or in impossibilities. These conflicts are detected by the process during the integration and/or the propagation steps because spatial integrity constraints are violated or because some updates have not been processed. The user is then compelled to

either solve these conflicts interactively or enhance the rule base and trigger anew the integration and propagation steps (cf. sections 10.4.2 and 10.4.3).

10.4.5 Checking database consistency

Once the updates have been propagated, several tests are performed to check the consistency of the whole derived database. These different tests consist in checking the topology of the database, the compliance with the metadata available on the derived database which have been updated. Conflicts possibly detected in this step are listed by the process and can be solved either interactively or by enhancing the integration and the propagation rule base by triggering anew the step described in section 10.4.4. The update which is the source of this conflict is thus rolled back.

In order to illustrate how the different steps of the updating mechanism interact, an overview of the process is presented in the following figure:

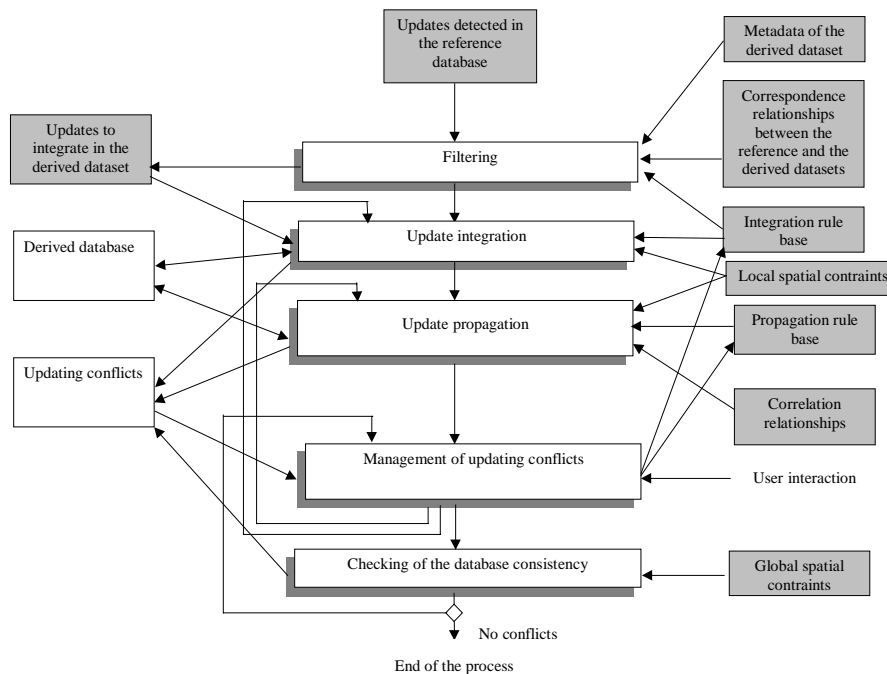


Figure 10.2 Overview of the process of update propagation.

In order to prove the feasibility of such an updating process, examples of strategies and of rules stemming from two experiments are presented in the next section.

10.5 EXPERIMENTS

The first experiment concerns the propagation of updates from BDTPOPO® (a 1:25,000 topographic database with 1-meter resolution produced by the IGN) to its derived cartographic database (TOP25®). The second one is about the updating of Route500®, a

geographic database stemming from a generalisation of the road network of BDCARTO® (a 1:100,000 database with 10-meter resolution also produced by the IGN).

These different tests which aim at validating the updating mechanism previously presented in this paper are in progress. The updates and the different correlation relationships have already been retrieved. Some steps of the propagation mechanism have been also defined and implemented. Some of these steps are thus reviewed in the following sections in order to illustrate the implementation of the rule base.

10.5.1 Update filtering

As detailed in section 10.4.1, this step allows the selection of the relevant updates which have to be integrated in the derived database. This selection relies on various types of information: the metadata describing the derived database, the derivation processes when they are formally defined and the correspondence relationships established between the reference and the derived data sets. In order to illustrate how the selection criteria are expressed, examples stemming from the experiment with BDCARTO® are presented below:

- Selection due to the metadata: As the geometrical resolution of Route500® is estimated to 100 meters, all displacements under 25 meters are not selected in BDCARTO®.
- Selection based on the derivation process: All elements of the secondary road network which have been updated in BDCARTO® are not selected because Route500® only includes a representation of the primary road network.

10.5.2 Update integration

In this section, we assume that the correlation relationships are established from the reference database to the derived one. For each type of evolution, algorithms involved in the integration of updates for the both experiments are defined. These algorithms are not completely detailed but principles of the method are provided. The detection of conflicts which may occur in this integration step is also presented below:

- Destruction: several cases have to be taken into account. Objects of the derived database involved in n-to-1 or n-to-m correspondence relationships are deleted only if the n objects of the reference database are also deleted. In all other cases (i.e. 1-to-1 and 1-to-n correspondence relationships), objects of the derived database are deleted.
- Creation: new objects of the reference database are created in the derived database by applying the derivation rules when they are defined or by requiring an interaction with the user.
- Semantic modification: Objects of the derived database involved in 1-to-1 or 1-to-n correspondence relationships are automatically updated. Otherwise, if the reference objects are not semantically updated in a similar way, the user is prompted to perform the updating interactively.
- Splitting: this type of update is automatically integrated if objects are referenced in 1-to-1 or 1-to-n correspondence relationships or if all objects of the reference database involved in another type of correspondence have undergone this modification for the same reason. Causes of the evolution have then to be

retrieved in order to determine if this splitting is due to an attribute modification or to an intersection with a new object. The splitting operation is then performed only if the attribute is also updated in the derived database or if the new object is also inserted. Nevertheless, this emphasises the role of derivation process in the definition of a strategy for the integration of updates. For instance, the derivation process between BDTOPO® and TOP25® is not formally determined, so the splitting point is located by using ratio on curvilinear abscissa for the reference object and then inserted in the derived objects. Finally, spatial integrity constraints have to be checked. Conversely, the derivation process between BDCARTO® and Route500® is completely defined. Objects involved in a relevant splitting operation are then derived by applying anew the derivation process. The consistency of the whole derived database is thus preserved.

- Merging: As before, this type of update is automatically integrated if objects are referenced in 1-to-1 or 1-to-n correspondence relationships or if all objects of the reference database involved in another type of correspondence have undergone this modification for the same reason. The integration is performed by merging the geometry of the derived objects when they are represented by contiguous lines or surfaces. The new attributes of the resulting object are then determined. Nevertheless, the integration is useless when the reference objects involved in the update are represented in the derived database by only one object.
- Aggregation: this type of update is processed with respect to the same requirements defined for the merging operations. However, the integration of these evolutions may be difficult to control so that the deletion of the old derived objects is first performed and the derivation rules are triggered on the reference objects involved in the aggregation to deduce the new version of the derived objects.
- Geometrical modification: this type of update may be processed if objects involved in the different correspondence relationships fulfil the requirements defined before. Otherwise, user is prompted to perform the updating interactively. With this type of update, the knowledge of the derivation process allows the determination of the most suitable algorithm. For instance, in BDCARTO®, derived objects involved in the geometrical modification are deleted and the update is integrated by applying the derivation rules to the reference object. This derivation process consists in the succession of a filtering algorithm as presented in Douglas and Peucker (1973) and of a smoothing algorithm. In BDTOPO®, the update is integrated by using a “morphing” algorithm, the distance between the derived and reference database is evaluated and the same difference between the updated reference object and the derived objects is performed.

For each type of evolution, the integration of update is possible when the updated objects are involved in 1-to-1 or 1-to-n correspondence relationships. With other types of correspondence (n-to-1 and n-to-m), the update can be automatically integrated only if the n reference objects are involved in the same modification. Otherwise the updating is performed by requiring an interaction with user. 1-to-0 correspondence relationships have to be checked by the user to ensure that the reference object is really not represented in the derived database.

10.5.3 Update propagation

In the experiment with BDTOPO®, some interesting correlation relationships have been detected. Indeed, as TOP25 is dedicated to cartography, some cartographic symbols which are derived from geographic objects have been added in it. For instance, bridges are symbolised on maps by short lines bordering the roads, and place names of rivers which are printed on maps follow the shape of rivers they name.

So, in this experiment, correlated objects have to be updated only when reference objects are involved in destruction or geometrical modification. For instance, in case of geometric modification, correlated objects have to be geometrically updated. However, propagating this type of update in TOP25® may be problematic because the correlated objects are produced by a derivation process and then checked interactively. So, the whole derivation process cannot be formally defined. Nevertheless, it is possible to define propagation strategies for specific geographical entities. For instance, the propagation of the updates performed on the bridges of the TOP25® consists in applying the same geometrical displacement to bridges' symbols. In addition, modifications performed on named objects may induce the update of their place names. As a place name is depending on several geographical objects, it is then necessary to first define if the modifications induce effects on it. If so, the place name is computed anew by applying the different propagation rules.

10.6 CONCLUSION AND OUTLOOKS

A generic updating tool for geographic databases which aims at automating the propagation of updates between databases with possible different levels of abstraction, representation and scale has been presented. This automation is provided by means of the implementation of a rule base which allows the control of the update integration and propagation. This rule base is fully upgradable and allows the consistent and complete updating of any kind of derived geographic database.

The implementation of this mechanism is in progress: the steps dealing with the management of conflicts have not been fully completed yet. Nevertheless, first results of the experiments presented in this paper have demonstrated its feasibility and about 80% of the updates have been consistently propagated. Once the implementation of the updating mechanism will be completed, the rate of automatic updating is expected to reach 90%.

10.7 REFERENCES

- Badard, T., 1998, Towards a generic updating tool for geographic databases. In *Proceedings of GIS/LIS'98, Annual Conference and Exposition*, Fort Worth, Texas, USA, pp. 352–363.
- Badard, T., 1999, On the automatic retrieval of updates in geographic databases based on geographic data matching tools. In *Proceedings of the 19th International Cartographic Conference (ICA'99)*, Ottawa, Canada, pp. 47–56.
- Bonnani, L., 1998, Etablissement de liens de corrélation dans un but de mise à jour des bases de données géographiques. Technical report IGN/DT/SR/980017, M. Sc. thesis in computer science, University of Paris-Dauphine, France, 65 pages.

- Douglas, D. and Peucker, D., 1973, Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. *The Canadian Cartographer*, **10(2)**, pp. 112–123.
- Egenhofer, M.J. and Herring, J.R., 1990, A mathematical framework for the definition of topological relationships. In *Proceedings of the 4th International Symposium on Spatial Data Handling (SDH'90)*, Vol.2, Zurich, Switzerland, pp. 803–813.
- Kilpeläinen, T., 1995, Updating Multiple Representation Geodata Bases By Incremental Generalization. In *Geo-Information-Systeme*, Wichmann, pp. 13–18.
- Laurini, R. and Thompson, D., 1992, *Fundamentals of spatial information systems*, (London: Academic Press).
- Lemarié, C. and Raynal, L., 1996, Geographic data matching: first investigations for a generic tool. In *Proceedings of GIS/LIS'96, Annual Conference and Exposition*, Denver, Colorado, USA, pp. 405–420.
- Lemarié, C. and Bucaille, O., 1998, Spécifications d'un module générique d'appariement de données géographiques. In *Actes du 11^{ème} congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'98)*, Clermont-Ferrand, France, pp. 397–406.
- Lemon, D. and Masters, E., 1997, The nature and management of positional relationships within spatial databases. In *Proceedings of the 2nd Annual Conference of Geocomputation'97*, University of Otago, New Zealand, pp. 15–23.
- Lupien, A.E. and Moreland, W.H., 1987, A general approach to map conflation. In *Proceedings of the ACSM/ASPRS Annual Convention & Exposition, Technical Papers, AutoCarto 8*, Baltimore, Maryland, USA, pp. 630–639.
- Saalfeld, A., 1988, Conflation – Automated map compilation. *International Journal of Geographical Information Systems (IJGIS)*, **2(3)**, pp. 217–228.
- Uitermark, H., van Oosterom, P., Mars, N. and Molenaar, M., 1998, Propagating updates: Finding Corresponding objects in a multi-source environment. In *Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98)*, Vancouver, BC, Canada, edited by Poiker, T.K., and Chrisman, N., (Burnaby: International Geographical Union), pp. 580–591.

4.3 Compléments

Cette section vise à apporter un certain nombre de compléments d'information sur le processus d'intégration et de propagation présenté dans la section précédente. Ils concernent le modèle de gestion sur lequel repose ce processus et la stratégie de résolution des conflits de mise à jour qui y est employée.

4.3.1 Modèle de gestion de l'intégration et de la propagation

Cette section s'attache à présenter le modèle de gestion sur lequel s'appuie le mécanisme d'intégration et de propagation établi dans la section 3.2. Défini dans [Badard et Lemarié, 2000], ce modèle exprimé dans le formalisme OMT [Rumbaugh et al., 1991] est fourni en Figure 4.1.

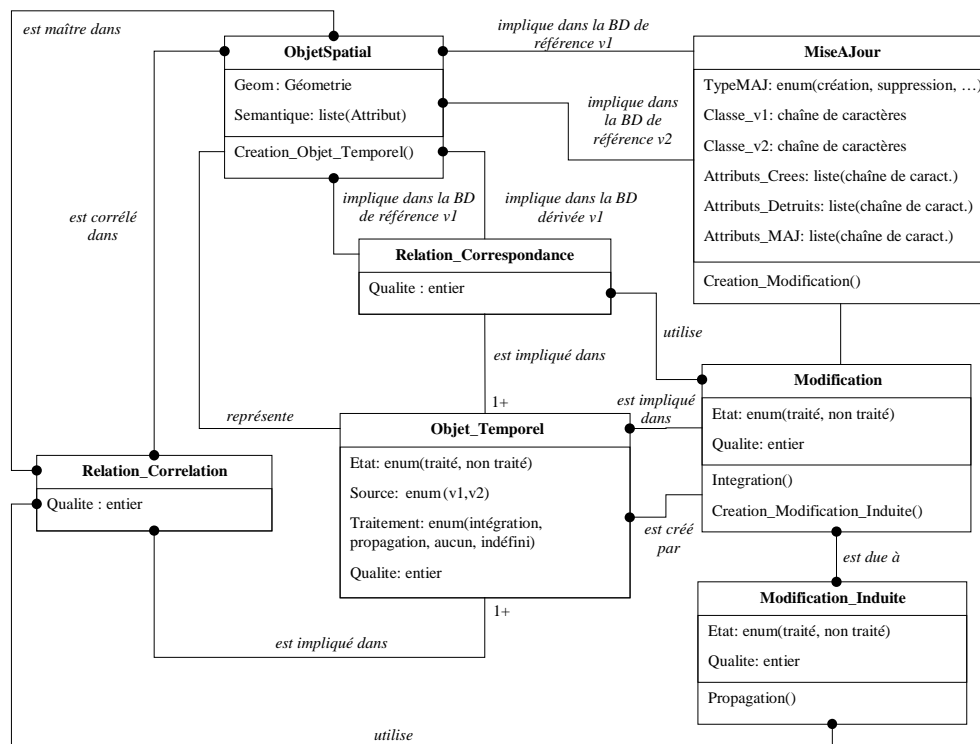


Figure 4.1 - Modèle de gestion de l'intégration et de la propagation des mises à jour

Le système comprend au départ, les jeux de données de référence dans leurs versions ancienne (*BD de référence v1*) et à jour (*BD de référence v2*), ainsi que la base dérivée ancienne (*BD dérivée v1*).

La classe **ObjetSpatial** permet de représenter tous les objets géographiques stockés dans les différentes bases de données. Chaque objet se compose d'une partie sémantique et d'une partie géométrique. Cette classe constitue en fait la classe mère du prototype de système de gestion de base de données géographiques GÉO₂., sur lequel nous avons menés nos tests et réalisés nos développements.

La classe **Relation_Correspondance** permet, comme son nom l'indique, le stockage des relations de correspondance extraites entre les versions anciennes du référentiel et de la base dérivée. De même, la classe **Relation_Correlation** est employée pour permettre l'expression des relations de corrélation extraites (hors celles de correspondances qui sont stockées via la classe **Relation_Correspondance**) entre les objets de la base de données dérivée mise en jeu (*BD dérivée v1*). L'attribut *Qualite*, porté par ces classes, permet d'exprimer la valeur du taux de confiance que l'on peut accorder à l'exactitude de la relation établie (cf. section 3.3.1.2.2, pour notion de qualité de lien d'appariement).

Afin de pouvoir gérer les différentes actualités des jeux de données présents dans le système et surtout afin de permettre l'expression de la relation temporelle de succession entre les entités dérivées dans leurs versions anciennes et mises à jour (instances de la classe **ObjetSpatial**), la classe **Objet_Temporel** a été définie.

Ainsi, à l'initialisation du processus de mise à jour, une instance de la classe **Objet_Temporel** est créée pour chaque objet de la base dérivée dans sa version ancienne. Son *Etat* est alors renseigné par la valeur « non traité » (cet attribut passera bien entendu à « traité » dès que l'objet aura été mis à jour). *Source* est égal à v1. La valeur de *Traitement* dépend du nombre et de la nature des relations qui lient cet objet. Par exemple, si l'objet n'est impliqué dans aucune relation de correspondance ou de corrélation, cet attribut prendra la valeur « aucun » (i.e., il ne subira pas d'opération de mise à jour).

La classe **MiseAJour** permet l'expression de toutes les évolutions extraites à partir de la comparaison des deux versions de la base de données de référence. Les mises à jour issues du processus d'extraction des évolutions sont donc toutes instances de cette classe.

L'attribut *TypeMAJ* permet de renseigner du type de mise à jour subie par les objets impliqués dans cette opération. Il prendra donc une des valeurs suivantes : création, destruction, modification sémantique, modification sémantique et géométrique, modification de classe, modification de classe et de la géométrie, modification géométrique, fusion, scission et agrégation. Les différentes versions des objets de la base de référence impliqués dans une opération de mise à jour sont liés au moyen de cette classe.

Les attributs *Classe_v1* et *Classe_v2*, portés par cette classe, permettent de stocker les noms des classes auxquelles appartiennent les objets impliqués dans la mise à jour. Les noms des attributs créés, détruits et mis à jour sont renseignés respectivement dans *Attributs_Creés*, *Attributs_Détruits*, et *Attributs_MAJ*. Cet ensemble d'attributs permet donc de transcrire les modifications de schéma intervenus dans la base de référence ou les changements de classification subis par certains objets de celle-ci.

Une fois que les mises à jour de la base de référence sont identifiées, c'est-à-dire que toutes les instances de la classe **MiseAJour** sont créées, la phase d'intégration et de propagation peut-être initiée.

Une mise à jour est intégrée dans la base de données dérivée au moyen d'une instance de la classe **Modification**. Une *Modification* n'implique qu'une et une seule *MiseAJour*. Elle est d'ailleurs créée à partir de cette dernière via l'utilisation de la méthode *Creation_Modification()*.

Afin de retrouver les objets de la base de données dérivée qu'elle doit mettre à jour (un ensemble d'*Objet_Temporel*), une *Modification* utilise les instances de la classe

Relation_Correspondance. En fonction du type de mise à jour (attribut *TypeMAJ* de la classe **MiseAJour**), de la cardinalité des liens de correspondance impliqués, de la valeur de l'attribut *Qualite* porté par ces liens et de la disponibilité éventuelle d'un processus de dérivation, l'intégration est déclenchée via la méthode *Integration()*. C'est donc à ce niveau là, que l'appel à la base de règles est réalisé.

Un ensemble de nouveaux objets de la classe **Objet_Temporel** (ainsi que les instances de la classe **ObjetSpatial** qu'ils représentent) est alors généré. Les liens entre les nouveaux *Objet_Temporel* et la *Modification* sont créés. On dispose ainsi de l'information de succession temporelle entre les objets de la base dérivée dans leur versions anciennes et mises à jour. L'attribut *Etat* de la *Modification* impliquée, est modifié (passage à la valeur « traité »)

La *Modification* étant intégrée, le processus vérifie alors, si les *Objet_Temporel* anciens sont impliqués dans des instances de la classe **Relation_Correlation** :

- Dans l'affirmative, une *Modification_Induite* est créée via l'utilisation de la méthode *Creation_Modification_Induite()*. L'application de cette mise à jour induite (via l'utilisation de la méthode *Propagation()*) se passe alors de façon analogue au processus décrit précédemment pour les objets de la classe **Modification**. In fine, l'attribut *Etat* des *Objet_Temporel* anciens, impliqués dans l'opération de mise à jour, est modifié (passage à la valeur « traité ») et on passe alors à l'opération de mise à jour suivante.
- Dans la négative, l'attribut *Etat* des *Objet_Temporel* anciens, impliqués dans l'opération de mise à jour, est modifié (passage à la valeur « traité ») et on passe à la *Modification* suivante.

Ce modèle permet donc de structurer facilement toute l'information nécessaire au processus de mise à jour. Il rend possible une intégration et une propagation aisée et surtout contrôlable des mises à jour dans les bases de données géographiques. De plus, il met en œuvre un système relativement simple de gestion d'historique qui permet d'assurer le lien de succession temporelle entre les objets des différentes bases mises en jeu. Ce système, du fait de l'utilisation lors des différentes phases d'extraction (mise à jour et lien de corrélation) de méthodes basées essentiellement sur la géométrie et la topologie, permet de s'absoudre de l'existence d'identifiants externes sur les objets.

De plus, toutes ces classes étant rendues persistantes dans le système, il permet de capitaliser toutes les informations utiles au processus de mise à jour et dont l'extraction peut être parfois coûteuses en temps de calcul. Ainsi, l'introduction de nouvelles évolutions dans le système ne nécessitera pas le ré-établissement des relations de correspondance et de corrélation. Ce modèle permet donc d'automatiser les opérations de mise à jour suivantes. Ainsi, il agit un peu, comme le modèle sous-jacent permettant la rétro-conception des systèmes d'information géographique afin d'assurer leur maintenance. Il préfigure de ce que pourrait être un modèle de gestion des évolutions, sur lequel s'appuierait une méthodologie de conception des système d'information géographique prenant en compte la notion de mise à jour, méthodologie de conception qui fait actuellement cruellement défaut, si l'on considère l'ensemble des difficultés auxquelles nous avons été confronté.

4.3.2 Précisions sur la gestion des conflits de mise à jour

Cette section s'attache à apporter des précisions et des éléments de réflexion sur la stratégie de résolution des conflits de mise à jour qui est employée au sein du processus d'intégration et de propagation.

Comme mentionné dans la section 4.2, cette stratégie se décompose en deux grandes étapes :

- Tout d'abord, chaque mise à jour détectée et filtrée de la base de référence est propagée de façon automatique, suivant un ordre de traitement défini dans la base de règles.
- Une fois que l'ensemble des mises à jour a été propagé, les conflits détectés au cours de l'étape précédente, sont mentionnés à l'utilisateur. Afin de les résoudre, il peut soit enrichir la base de règles et relancer une phase de propagation automatique, soit les corriger de façon interactive (en s'appuyant sur des outils dédiés).

Ceci fait apparaître le rôle critique joué par le choix de l'ordre dans lequel on va propager les mises à jour qui ont été détectées et sélectionnées dans la base de référence. En effet, au cours de l'étape de propagation, la détection des conflits va se baser sur un certain nombre de tests concernant non seulement les données anciennes de la base dérivée qui n'évolueront pas ou qui n'ont pas encore été mises à jour, mais également les objets qui ont été déjà modifiés dans celle-ci. Par exemple, pour des données de type réseau, on va vérifier les raccordements aux réseaux ancien et nouveau, les intersections éventuelles avec les réseaux ancien et nouveau.

Il se peut donc qu'au cours de la propagation, des mises à jour soient détectées comme conflictuelles, car tout simplement, un certain nombre d'évolutions liées (i.e., dont l'application conduirait à l'établissement d'une mise à jour correcte) n'ont pas encore été appliquées. Même si, en l'état actuel de sa mise en œuvre, la base de règles permet déjà un regroupement des mises à jour qui dépendent les unes des autres, tous les cas sont difficiles à prendre en compte, avant propagation, avec un tel système de règles préétablies. Aussi, si l'on veut atteindre un taux élevé de propagation entièrement automatique, devra-t-on rendre le mécanisme plus flexible, autonome et intelligent, en lui permettant notamment d'opérer par lui-même, cette étape de regroupement et de sélection des mises à jour à appliquer.

Ces notions de but à atteindre et d'autonomie sont au cœur des systèmes basés sur les concepts de contraintes, tels que mis en œuvre pour le processus de généralisation par [Ruas, 1999], ou de multi-agents [Ferber, 1995 ; Demazeau, 1995]. Aussi, la mise en œuvre de telles techniques au sein du mécanisme de mise à jour, pourrait-elle, très certainement, permettre d'améliorer ce manque de flexibilité et d'intelligence du processus d'intégration et de propagation.

4.4 Résultats des cas d'étude

Les deux sections suivantes présentent, de façon très synthétique, les résultats produits par le mécanisme d'intégration et de propagation des mises à jour, sur les deux cas d'étude qui ont été entrepris. En effet, les statistiques détaillées sur les nombres d'objets impliqués dans les différents bases mises en jeu, ne présentant que peu

d'intérêt, nous nous attacherons plutôt à donner et commenter les taux de réussite du processus.

4.4.1 Propagation des mises à jour de la BDCarto vers Route500

A l'issue de l'étape de filtrage, basée sur le détournement du processus de sélection de Route500, seulement 1533 mises à jour, sur les 8467 extraites de la BDCarto, ont été retenues, soit environ 18 %. Ce taux de mises à jour sélectionnées est relativement cohérent avec le rapport du nombre d'objets contenus dans Route500 et de celui de la BDCarto (22 %).

La totalité de ces mises à jour est ensuite passée par les phases d'intégration et de propagation du processus. Au cours de ceux-ci, 352 opérations de mise à jour ont été jugées comme conflictuelles, soit 23 % des évolutions issues de la phase de filtrage. La carte des conflits détectés est donnée dans la Figure 4.2.

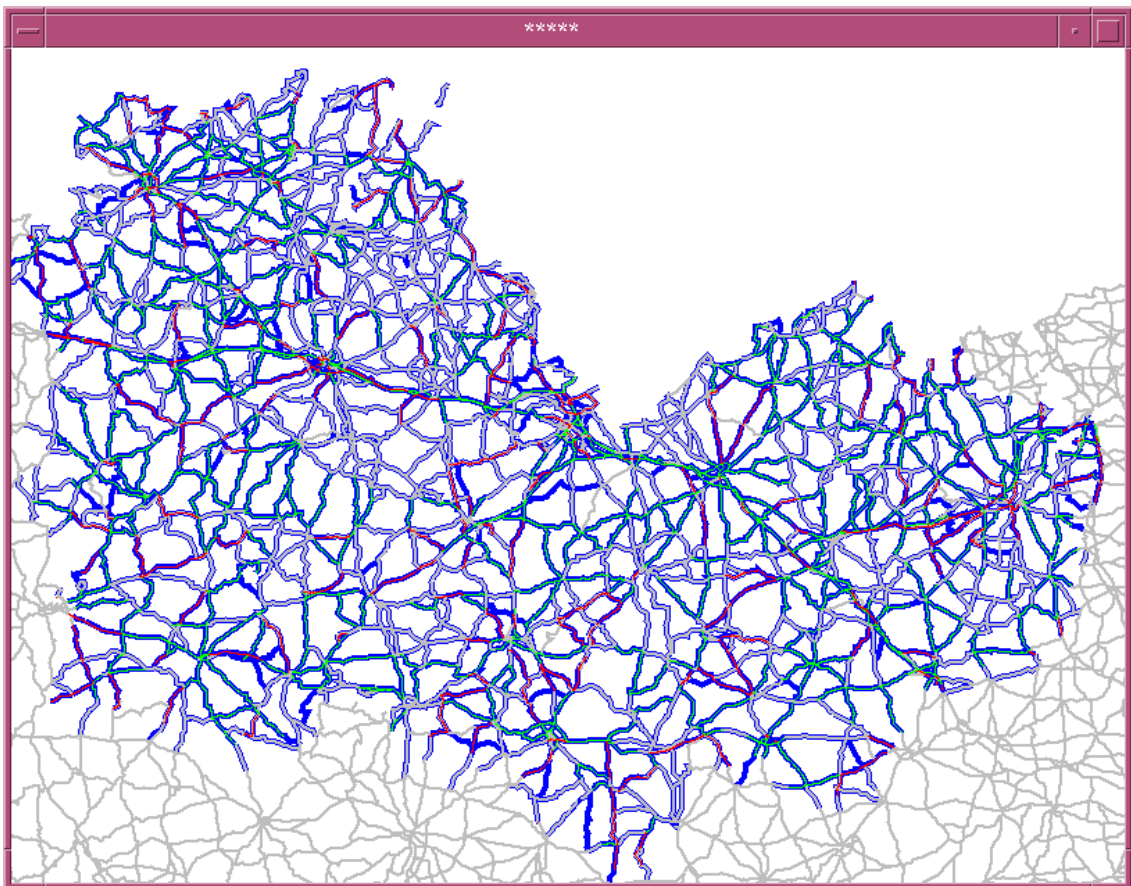


Figure 4.2 – Carte des conflits d'intégration des mises à jour BDCarto dans Route500

Légende :

- BLEU : Objets Route500 dans leur version à jour produite par l'IGN.
- ROUGE : Conflits de mise à jour détectés.
- VERT : Objets mis à jour sans conflits.
- GRIS : Objets Route500 dans leur version ancienne.

Afin d'aider à la reprise de ces conflits, une interface dédiée a été développée. Elle est illustrée par la Figure 4.3.

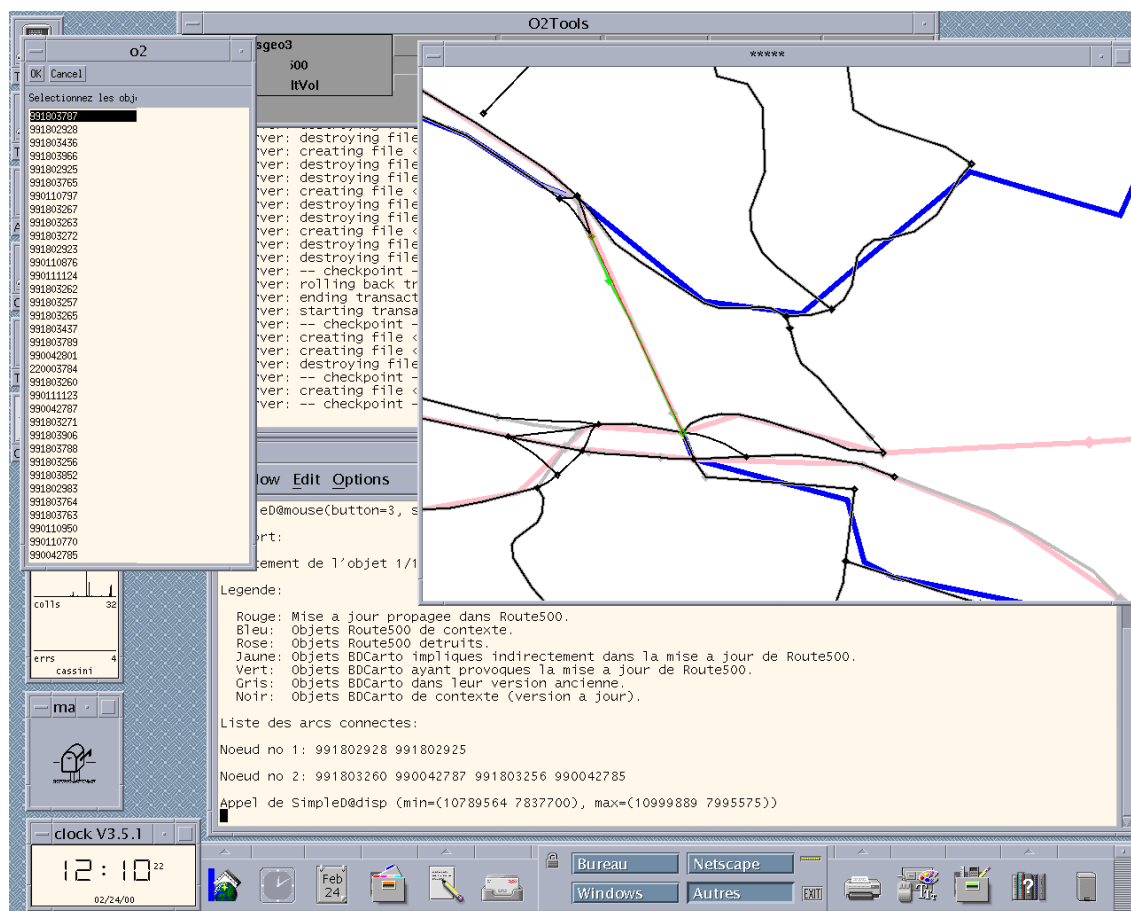


Figure 4.3 – Interface de résolution des conflits de mise à jour

Comme le montre la figure ci-dessus (la légende est fournie à l'intérieur de celle-ci), cette interface propose à l'utilisateur une liste des arcs BDCarto, les plus probables, qu'il faudrait rajouter ou enlever à la sélection déjà opérée, pour permettre une propagation de la mise à jour qui soit cohérente. Elle permet donc de guider efficacement un opérateur dans la phase de reprise et diminue de façon très significative les temps passés à la résolution des conflits.

Le conflit illustré dans la Figure 4.3, provient d'un problème de sous sélection inhérent au processus de dérivation. En effet, nous l'avons mentionné (cf. section 2.6.1), la réalisation de Route500 nécessite une part importante de reprises manuelles à l'issue de la phase automatique. Ceci est justement dû au processus de sélection utilisé dans la production de Route500. Il a, en effet, tendance à sur ou sous sélectionner les objets de la BDCarto qui interviendront ensuite dans la phase de généralisation, tout en n'assurant pas la continuité et la logique de desserte du réseau sélectionné. Toutes les mises à jour en conflit résultent de ce problème.

En conclusion de ce test, nous pouvons donc constater que 77% des mises à jour ont pu être propagées automatiquement, et l'ont été correctement au regard de comparaisons effectuées avec la version à jour produite par l'IGN. Les 23 % d'évolutions en conflit, détectées par le mécanisme, sont toutes le résultat des

problèmes inhérents au processus de sélection opéré dans la dérivation de Route500. L'enrichissement de la base de règles n'étant pas alors possible, une interface, guidant la reprise de ceux-ci, a été développée et permet de les corriger rapidement. Les conflits étant identifiés et localisés par le mécanisme (ce qui n'est pas le cas lorsque l'on dérive le jeu entier), le temps, et donc le coût, de la phase de reprise manuelle est très considérablement réduit.

4.4.2 Propagation des mises à jour de la BDTopo vers BDTopo25

A l'issue de l'étape de filtrage, basée sur le détournement du processus de sélection de la BDTopo25, l'ensemble des mises à jour extraites de la BDTopo ont été jugées pertinentes. Ce taux de mises à jour sélectionnées est relativement cohérent avec le rapport du nombre d'objets contenus dans la BDTopo25 et de celui de la BDTopo (91 %).

La totalité de ces mises à jour est ensuite passée par les phases d'intégration et de propagation du processus. Au cours de ceux-ci, 435 opérations de mise à jour ont été jugées comme conflictuelles, soit environ 9 % des évolutions issues de la phase de filtrage. Les figures ci-dessous présentent les résultats obtenus par le mécanisme de mise à jour sur deux extraits de la zone d'étude.



Figure 4.4 - Résultats du test de propagation des mises à jour de la BDTopo sur la BDTopo25 (extrait 1)

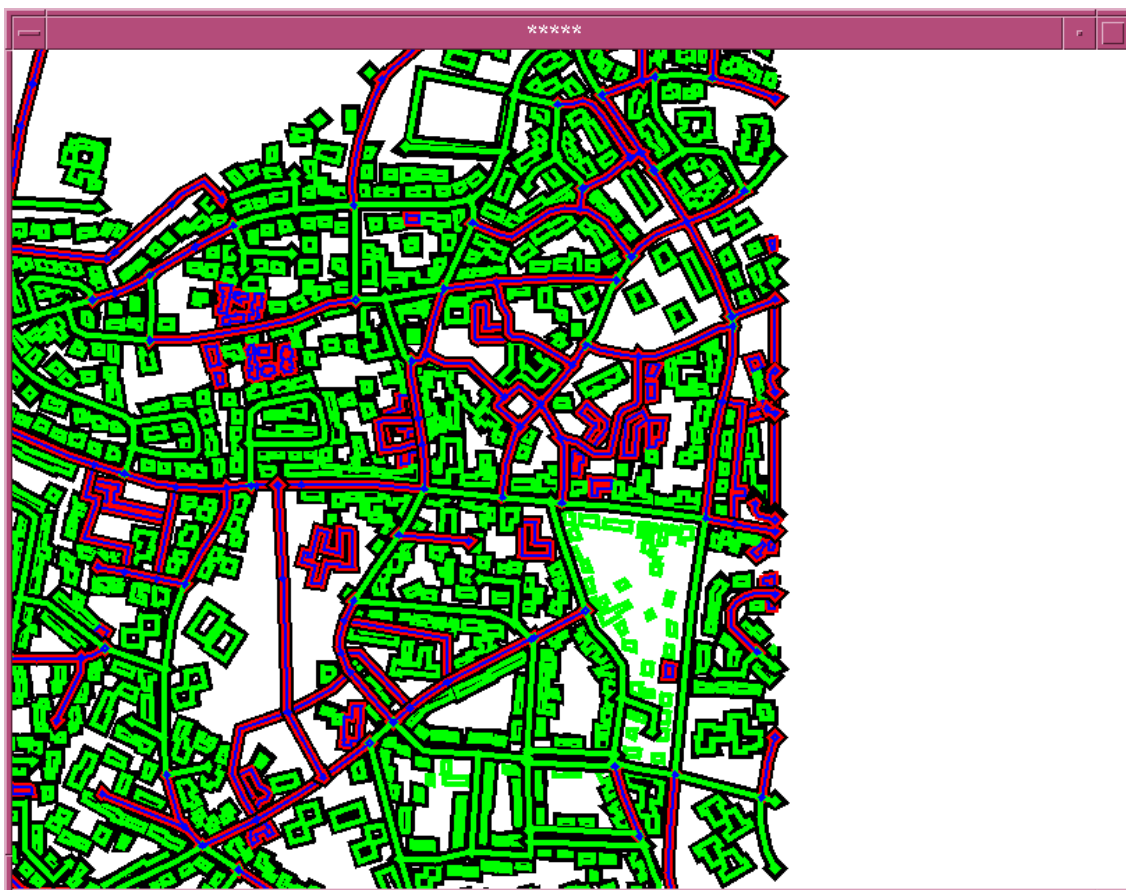


Figure 4.5 - Résultats du test de propagation des mises à jour de la BDTopo sur la BDTopo25 (extrait 2)

Légende :

- NOIR : Objets de la BDTopo25 à jour tels que produits par l'IGN.
- ROUGE : Objets TOP25V2 à jour issus de la propagation.
- VERT : Objets TOP25V1.
- BLEU : Objets BDTopo dans leur version à jour (BDTOPOV2), intervenants dans une mise a jour.

En fait, faute de temps, nous n'avons pas pu traiter l'ensemble des classes de la BDTopo. Nous nous sommes, en effet, intéressés exclusivement aux thèmes routier, bâti, ferré et monuments de cette base, ce qui a, rappelons le, tout de même permis de propager 91 % des 4548 mises à jour extraites. Sur les 435 opérations de mise à jour détectées comme conflictuelles, 398 (92 %) mettent en jeu des classes d'objets que nous n'avons pas pris en compte. Les 8 % restants (37 mises à jour) correspondent à des cas complexes et particuliers de propagation (impliquant notamment plusieurs objets TOP25V1 liés) que nous n'avons, là aussi, pas eu le temps de traiter, en incluant les traitements dédiés dans la base de règles.

Ainsi, si l'on exclut ces cas, l'ensemble des mises à jour détectés et filtrés a été intégré et propagé avec succès dans la BDTopo25. Les mises à jour non traitées sont néanmoins identifiées et localisées (objets en vert, rayé bleu, sur les extraits) par le mécanisme et pourront être reprise en incluant les traitements spécifiques dans la base de règles.

La comparaison des deux jeux BDTopo25 à jour, l'un issu du mécanisme de propagation et l'autre produit par l'IGN, laisse apparaître quelques décalages de géométrie, notamment sur le réseau routier, tels que illustrés par la Figure 4.6.



Figure 4.6 - Illustration des décalages de géométrie entre versions à jour de la BDTopo25

Ces décalages proviennent tous de la phase de reprise manuelle opérée lors de la dérivation de la BDTopo25 à partir de la BDTopo. En effet, afin d'améliorer la lisibilité de la carte, les objets, dont les symbolisations se recouvrent, sont soit déplacés (quitte, d'ailleurs, à ne pas préserver la topologie de la base...), soit éliminés au cours de cette phase de reprise. Faute de temps, là aussi, nous n'avons pu développer de mécanisme permettant la détection de ces cas de figure particuliers (moins d'une dizaine sur l'étendue de la zone traitée). Néanmoins, la prise en compte des largeurs de symboles dans l'étape finale de contrôle de la base mise à jour, permettrait de détecter ces configurations et donc via le développement d'une interface de reprise dédiée, permettre une correction rapide de celles-ci.

Chapitre 5

Conclusion et perspectives

5.1 Conclusion

Afin de permettre une intégration des mises à jour d'une base de référence et une propagation des effets de celles-ci dans les bases de données géographiques utilisateur ou dérivées, qui soit la plus automatique possible et qui n'entraîne pas de dégradations de la cohérence et des acquis, nous avons tout d'abord établi une taxonomie complète des problèmes entravant la maintenance de ces systèmes multi-représentations.

Nous nous sommes, ensuite, attachés à considérer les solutions et travaux qui ont été développés jusqu'à maintenant et qui visent à aider à la mise à jour des bases de données géographiques. Nous nous sommes tout particulièrement efforcés à mettre en exergue les apports qu'elles fournissent pour atteindre le but qui nous nous étions fixé.

S'appuyant sur la taxonomie que nous avons établi et sur les apports fournis par les solutions actuellement mises en œuvre, nous avons défini et développé un mécanisme qui permet un niveau élevé d'automatisation de l'opération de mise à jour des systèmes d'information géographique multi-représentations et qui minimise les risques de perte d'information ou d'obtention d'états incohérents dans de tels systèmes.

Ce mécanisme qui s'étend de l'extraction de l'information d'évolution dans une base de données de référence jusqu'à l'intégration et la propagation de ces mises à jour dans les systèmes utilisateur ou dérivés, se décompose en trois sous mécanismes.

1. Le premier sous mécanisme que nous avons, là aussi, entièrement défini et mis au point, permet de détecter et d'extraire automatiquement les mises à jour intervenues dans une base de référence par comparaison de deux actualités de celles-ci. S'appuyant sur des méthodes d'appariement de données géographiques, ce mécanisme ne nécessite aucunement, l'existence ni même l'utilisation de systèmes d'identification ou de gestion de l'historique dans les bases de données mises en jeu. Ce choix vient du fait qu'à l'heure actuelle les identifiants sont mal ou non gérés. Nous nous sommes donc basés sur les caractères géométriques et sémantiques pour progressivement apparier les objets afin de détecter les évolutions. L'information qu'il fournit et qui concerne la nature des évolutions intervenues, est proche des modifications opérées et opérables dans les systèmes d'information géographique. Il constitue un préalable indispensable à une intégration, plus aisée et maîtrisable, des évolutions dans les bases de données utilisateur ou dérivées.

2. Le deuxième sous mécanisme, que nous avons mis au point, rend possible la détection et l'établissement de toutes les relations nécessaires à une intégration et à une propagation, sans perte d'information ou obtention d'états incohérents, des évolutions dans les systèmes. Cette étape permet donc d'identifier les relations de correspondances entre la base de référence et la base dérivée ou utilisateur et les relations de corrélation qui existent au niveau de la base dérivée. Ce mécanisme s'appuie partiellement sur les mêmes méthodes d'appariement de données géographiques que celles que nous avons utilisées précédemment. Il permet, en quelque sorte d'aider à la ré-ingénierie des systèmes d'information utilisateur ou dérivés afin de les rendre aptes à intégrer les évolutions intervenues dans une base de référence et sur lesquels ils s'appuient.
3. Le dernier sous mécanisme que nous avons également développé, constitue le cœur du processus de mise à jour. Utilisant l'information d'évolution extraite de la base de référence et les relations précédemment établies, ce mécanisme permet d'assurer la cohérence et l'automatisation de l'intégration et de la propagation des mises à jour dans les systèmes d'information utilisateur ou dérivés. Ce mécanisme met en œuvre une base de règles qui définit la stratégie de mise à jour à adopter en fonction des systèmes cibles et ainsi, rend possible le déclenchement des traitements appropriés pour mener à bien l'opération de mise à jour. Il permet, de plus, la détection des conflits de mise à jour qui peuvent intervenir et propose un certain nombre d'outils facilitant la reprise de ceux-ci.

Le mécanisme de mise à jour que nous avons défini et développé permet, de surcroît, la capitalisation de toutes les informations qui lui sont utiles. Ainsi, l'application de nouvelles mises à jour nécessitera moins de calculs, qui peuvent parfois être relativement coûteux en temps machine, et permettra donc d'automatiser encore les opérations de mise à jour suivantes.

Nous avons finalement mis en œuvre ce mécanisme, afin de le tester, dans le cadre de deux cas d'études faisant intervenir des bases de données produites par l'Institut Géographique National et présentant des niveaux d'abstraction, d'échelle ou de représentation différents. Les résultats que nous avons obtenus ont permis de le valider dans le cas où les processus de dérivation sont au moins partiellement connus et maîtrisés. Il resterait, néanmoins, à le tester sur des jeux de données faisant intervenir des processus plus complexes voir totalement inconnus, afin d'achever sa validation.

Enfin, afin de répondre aux problèmes d'isolation et d'accès à l'information d'évolution de la donnée géographique et en quelque sorte, de mettre à disposition des utilisateurs une partie de ces travaux de recherche, nous avons également défini de nouveaux modes de livraison de l'information de mise à jour (notions de lots différentiels et d'évolution). Ceux-ci font déjà actuellement l'objet d'expérimentations auprès d'utilisateurs tant internes qu'externes à l'IGN. Leur industrialisation dans le cadre d'un projet engagé par l'Institut, visant à mettre à disposition et à diffuser ses différentes productions de données numériques sur les réseaux tels qu'Internet, est très fortement envisagée.

5.2 Perspectives

Les perspectives offertes par ce travail de recherche sont diverses et multiples. Elles concernent non seulement l'amélioration du fonctionnement et des résultats du mécanisme que nous avons développé mais aussi la mise à disposition de ces travaux de recherche auprès des utilisateurs, dont l'IGN (en tant qu'utilisateur de ses propres données). Ainsi la mise à jour de la BDTopo et de ces produits dérivés (pour l'instant la BDTopo25, peut-être plus tard la BDCarto) devient tout à fait envisageable en production. De façon plus générale, de nouvelles problématiques liées à la mise à jour des systèmes d'information géographique peuvent également être adressées grâce à celui-ci.

Déjà introduite dans les différents compléments que nous avons fournis dans ce mémoire, une voie d'amélioration du fonctionnement et donc des résultats qu'il produit serait l'introduction de techniques d'apprentissage ou basées sur des systèmes mixtes de règles et de contraintes, voire même de multi-agents, tant dans le processus d'appariement de données géographiques que dans le mécanisme d'intégration et de propagation des mises à jour. La mise en œuvre de telles techniques permettrait, très certainement, d'améliorer le manque de flexibilité, d'autonomie et d'intelligence actuel, que nous avons mis en exergue dans le fonctionnement de ces deux processus. Le degré d'automatisation de l'opération de mise à jour qu'offre déjà, le mécanisme que nous avons défini et développé, s'en trouverait à n'en pas douter accru.

L'emploi de logiques de défaut, voire même de logique floue, au sein du mécanisme d'appariement de données géographiques, peut également améliorer la qualité des résultats qu'il produit. Ces logiques permettraient, en effet, non seulement de rendre son fonctionnement moins « binaire » mais aussi d'estimer le taux de confiance que l'on peut accorder aux liens de correspondance qu'il établit. Si leur introduction pose, néanmoins, le problème de la gestion de la propagation et de l'estimation de cette information au cours du processus d'appariement, cette information peut s'avérer cruciale pour toutes les applications s'appuyant sur la mise en correspondance de jeux de données géographiques.

Les nouveaux modes de livraison de l'information de mise à jour, définis et mis au point au cours de ce travail de thèse, et notamment la notion de lots d'évolution, basés sur XML (eXtensible Markup Language) permettent clairement d'entrevoir des applications de mise à jour des systèmes d'information géographique au travers de réseaux tels qu'Internet. De telles applications ouvrent ainsi de nouvelles perspectives de recherche, liées à l'intégration et à la propagation des évolutions dans les bases de données géographiques réparties. L'exploration de ces nouvelles thématiques vient d'être engagée au laboratoire COGIT et les résultats qu'elle produira, s'intégreront, à n'en pas douter, dans les réflexions en cours au sein de l'Institut, au sujet de l'unification des différentes bases de données géographiques qu'il produit et de la mise à jour délocalisée d'une telle base unifiée.

Même si des telles applications ne sont pas encore à la portée de tous les utilisateurs, nous avons déjà réalisé le portage d'un certain nombre d'algorithmes d'appariement de données géographiques sur des logiciels SIG du marché. Ces développements peuvent donc les aider à intégrer dans leurs systèmes, les évolutions que le producteur leur fournira. De même, la taxonomie des problèmes entravant l'intégration et la propagation des évolutions dans les bases de données géographiques, que nous avons dressée dans ce mémoire, peut également aider les utilisateurs à

concevoir de nouveaux systèmes d'information géographique, offrant une maintenance plus aisée. En effet, elle offre à ceux-ci, une sorte de guide permettant la mise en œuvre de tels systèmes.

La définition d'une véritable méthodologie de conception de systèmes d'information géographique, permettant une automatisation de l'opération de mise à jour et se basant sur cette taxonomie, ainsi que sur l'ensemble des outils développés au cours de ce travail de thèse, fait actuellement l'objet de recherches au sein du laboratoire COGIT (thèse de Hakima Kadri-Dahmani). Cette méthodologie de conception devra, de plus, permettre l'expression des modèles conceptuels des systèmes, au travers d'un formalisme dédié à la donnée géographique et intégrant la notion de temps. Les apports constitués par des formalismes comme MADS [Parent et al., 1998] ou Perceptory [Bédard, 1999], ce dernier s'appuyant sur UML [Booch et al., 1997] et offrant ainsi des possibilités d'extensions accrues, seront donc à considérer pour atteindre cet objectif. La génération automatique de ces modèles dans des langages pivots, afin de permettre l'implantation de ceux-ci dans des plate-formes du marché (i.e. accessibles aux utilisateurs), est également envisagée.

Bibliographie

- [**Abbas, 1994**] I. Abbas. Base de données vectorielles et erreur cartographique : problèmes posés par le contrôle ponctuel - Une méthode alternative fondée sur la distance de Hausdorff : le contrôle linéaire. Thèse de doctorat de l'université de Paris XII, Paris, France, 10 juin 1994.
- [**Abrantes, 1996**] G. Abrantes. The object-oriented approach to the specification of geographic applications. In Proceedings of the 2nd Joint European Conference (JEC'96), Rumor, McMillan et Otten (Eds.), IOS Press, Barcelona, Spain, 1996, pp. 48–57.
- [**AFNOR, 1992**] AFNOR. Echanges de Données Informatisés dans le Cadre de l'Information Géographique (EDIGÉO). AFNOR, 1992,. ISSN 0335-3931.
- [**Amer-Yahia et al., 1996**] S. Amer-Yahia, P. Brèche, C. Souza dos Santos. Object Views and Updates. Actes des 12^{èmes} journées Bases de Données Avancées (BDA'96), Cassis, France, 27-30 Août 1996, 19 pages.
- [**Badard, 1998a**] T. Badard. Mise à jour générique des bases de données géographiques : Panorama des problèmes et premières investigations. Actes des 1ères journées Ré-ingénierie des Systèmes d'Information (RSI'98), Inforsid (Eds.), Lyon, France, 1-2 Avril, 1999, pp. 109–118.
- [**Badard, 1998b**] T. Badard. Premières pistes pour une mise à jour générique des bases de données géographiques. Bulletin d'information de l'IGN, n° 69 (1998/2), IGN, Paris, France, 1998, pp. 25–32.
- [**Badard, 1998c**] Extraction des mises à jour dans les BDG – De l'utilisation de méthodes d'appariement. Revue internationale de géomatique, vol. 8, n° 1-2, Hermès, Paris, 1998, pp. 121–147.
- [**Badard, 1998d**] T. Badard. Towards a generic updating tool for geographic databases. In Proceedings of GIS/LIS'98, Annual Conference and Exposition, AAG, ACSM, ASPRS, GITA, URISA (Eds.), Fort Worth, Texas, USA, November 1998, pp. 352–363.
- [**Badard, 1999**] T. Badard. On the automatic retrieval of updates in geographic databases based on geographic data matching tools. In Proceedings of the 19th International Cartographic Conference (Ottawa'99), ICA/ACI (Eds.), Ottawa, Ontario, Canada, 1999, pp. 47–56.
- [**Badard et Lemarié, 1999**] T. Badard et C. Lemarié. Propagating updates between geographic databases with different scales. Chapter 10 of Innovations in GIS VII: GeoComputation, Atkinson, P. and Martin, D. (Eds.), Taylor and Francis, London, 1999, 12 pages.
- [**Badard et Lemarié, 2000**] T. Badard et C. Lemarié. Propagating updates between geographic databases. In submission to the International Journal of Geographical Information Science (IJGIS), Taylor and Francis, London, 2000, 12 pages.
- [**Badard et Richard, 2000**] T. Badard et D. Richard. Using XML for the exchange of updating information between geographical information systems. To appear in the Computers, Environment and Urban Systems Journal (CEUS), Elsevier Science, Amsterdam, 2000, 12 pages.
- [**Bancilhon et al., 1992**] F. Bancilhon, C. Delobel et P. Kanellakis (Eds.). Building an Object-Oriented Database System – The Story of O2. Morgan Kaufman Publishers, San Mateo, California, 1992, ISBN 1-55860-169-4.
- [**Bédard et al., 1997**] Y. Bédard, Y. van Chestein, G. Poupart. Actualisation des données à référence spatiale (volets échange et intégration). Centre de Recherche en Géomatique, Université Laval, Québec, Canada, 1997, 54 pages.
- [**Bédard, 1999**] Y. Bédard. Visual modelling of spatial databases : towards spatial PVL and UML. Geomatica, vol. 53, no. 2, 1999, pp. 169–186.
- [**Bel Hadj Ali, 1997**] A. Bel Hadj Ali. Appariement géométrique des objets géographiques et étude des indicateurs de qualité. Mémoire de DEA Sciences de l'Information Géographique, Université de Marne la Vallée, Marne la Vallée, France, 18 Septembre 1997, 109 pages.
- [**Bauzer-Meideros et Jomier, 1994**] C. Bauzer-Medeiros et G. Jomier. Using versions in GIS. In Proceedings of the Databases and Expert Systems Applications Conference (DEXA'94), Athens, Greece, 1994.

- [**Bishr, 1998**] Y. Bishr. Overcoming the semantic and other barriers to GIS interoperability. *International Journal of Geographical Information Science (IJGIS)*, 12(4), Taylor and Francis, London, 1998, pp. 299–314.
- [**Bonnani, 1998**] L. Bonnani. Etablissement de liens de corrélation dans un but de mise à jour des bases de données géographiques. Rapport technique IGN/DT/SR/980017, Mémoire de DEA en informatique, Université Paris-Dauphine, France, 1998, 65 pages.
- [**Bonin, 1998**] O. Bonin. Attribute uncertainty propagation in vector geographic information systems : sensitivity analysis. In *Proceedings of SSDBM'98*, M. Rafanelli et M. Jarke (Eds.), IEEE/Computer Society, 1998, pp. 254–259.
- [**Booch et al., 1997**] G. Booch, J. Rumbaugh et I. Jacobson. *Unified Modeling Language user guide*. Addison-Wesley, 1997. URL : <http://www.rational.com/uml/references/docset.html>.
- [**Brooker, 1995**] P. Brooker. Development of a Multiresolutional Integrated Land Cover Database for Scotland and the Application of GIS in Predictive Spatial Modeling. In *Proceedings of the ESRI User Conference*, 1995.
- [**Bouillé, 1987**] F. Bouillé. A survey on the HBDS methodology applied to cartography and land planning. In *Proceedings of EuroCarto VI*, Brno, 1987, pp. 20–28.
- [**Brassel et Weibel, 1988**] K. Brassel et R. Weibel. A review and conceptual framework of automated map generalization. *International Journal on Geographical Information Systems (IJGIS)*, Vol. 2, Num. 3, Taylor and Francis, London, 1988, pp. 229–244.
- [**Brugger et al., 1989**] B. Brugger, R. Barrera, A. Frank, K. Beard et M. Ehlers. Research Topic on Multiple Representations. Workshop on Multiple Representations Initiative 3, National Center for Geographic Information and Analysis (NCGIA), 1989, pp. 53–67.
- [**Bucaille, 1996**] O. Bucaille. Mise à jour d'une base de données géographiques – Exemple de la BDCarto®. Mémoire d'ingénieur CNAM, France, 1996, 119 pages.
- [**Cellary et Jomier, 1990**] W. Cellary et G. Jomier. Consistency of versions in object oriented databases. In *Proceedings of the 16th Very Large Database Conference (VLDB'90)*, Brisbane, Australia, 1990, pp. 432–441.
- [**CEN, 1996**] CEN/TC 287. Geographic Information – Data Description – Transfer. Draft. CEN/TC 287, 1996.
- [**CEN, 1997**] CEN/TC 287. ENV 12160 : “Geographic information – Data description – Spatial Schema”, CEN/TC 287, 1997.
- [**CEN, 1999**] CEN/TC 287. ENV 12657 : “Geographic information – Data description – Metadata”, Rev 165, CEN/TC 287, April 1999.
- [**Cheyland et al., 1994**] J.P. Cheyland, S. Lardon, H. Mathian, L. Sanders. Les problématiques liées au temps dans les SIG. *Revue internationale de géomatique*, vol. 4, no 3-4, 1994, pp. 287–305.
- [**Claramunt et al., 1994**] C. Claramunt, M.H. de Sède, R. Prélaz-Droux, L. Vidale. Sémantique et logique spatio-temporelles. *Revue internationale de géomatique*, vol. 4, no 2, 1994, pp. 165–180.
- [**Cobb et al., 1998**] M. A. Cobb, F. E. Petry et K. B. Shaw. Uncertainty issues of conflation in a distributed environment. In *Proceedings of GIS/LIS'98, Annual Conference and Exposition*, Fort Worth, Texas, USA, 1998, pp. 436–448.
- [**Cordier et Siegel, 1994**] M. O. Cordier et P. Siegel. Prioritized Transitions for Updates. IRISA, publication interne no 884, Novembre 1994, 13 pages.
- [**David, 1991**] B. David. Modélisation, représentation et gestion d'information géographique. Mémoire de thèse de doctorat en Informatique de l'Université de Paris 6, Paris, France, 1991, 215 pages.
- [**David et al., 1993a**] B. David, L. Raynal et G. Schorter (1993 a) Evaluation of the OO approach for Geographical Applications. Technical Report of the AMUSING Esprit Project, num. 6881, 1993.
- [**David et al., 1993b**] B. David, L. Raynal, G. Schorter et V. Mansart. Why objects in a geographical DBMS ?. In D. Abel and B. Chin Ooi. (eds.), *Advances in Spatial Databases*, LNCS 692, Springer-Verlag, Berlin Heidelberg New York, 1993, pp. 264–276.
- [**David et al., 1993c**] B. David, L. Raynal, G. Schorter et V. Mansart. Géo₂ : modélisation objet de données. *Revue internationale de géomatique*, Hermès, vol. 3, no. 3, 1993, pp. 265–281.
- [**David et Fasquel, 1997**] B. David et P. Fasquel. Qualité d'une base de données géographiques : concepts et terminologie. *Bulletin d'information de l'IGN*, n° 67, IGN, Saint-Mandé, France.

- [**Dell'Erba et Libourel, 1997**] E. Dell'Erba et T. Libourel. Evolution d'entités géoréférencées. Actes des 13^{èmes} journées Bases de Données Avancées (BDA'97), Grenoble, France, Septembre 1997, pp. 89–110.
- [**Demazeau, 1995**] Y. Demazeau. From interactions to collective behaviour in agent-based systems. In Proceedings of the 1st European Conference on Cognitive Sciences, Saint Malo, France, 1995.
- [**Demirkesen et Schaffrin, 1996**] A. Demirkesen et B. Schaffrin. Map Conflation : Spatial point data merging and transformation, In Proceedings of GIS/LIS'96, Annual Conference and Exposition. Denver, Colorado, USA, November 19-21, 1996, pp. 393–404.
- [**Denègre et Salgé, 1996**] J. Denègre et F. Salgé. Les systèmes d'information géographique. Collection Que sais-je ?, no 3122, Presses Universitaires de France.
- [**Devoegele, 1997**] T. Devoegele. Processus d'intégration et d'appariement de bases de données géographiques – Application à une base de données routières multi-échelles. Mémoire de thèse de doctorat en Méthodes Informatiques de l'Université de Versailles, France, 12 décembre 1997, 207 pages.
- [**Devoegele et al., 1998**] T. Devoegele, C. Parent, et S. Spaccapietra. On spatial database integration. International Journal of Geographical Information Science (IJGIS), 12(4), P. Fisher, M. Armstrong et B. Lees (Eds.), June 1998, Taylor and Francis, London, pp. 335–352.
- [**Douglas et Peucker, 1973**] D. Douglas. et D. Peucker. Algorithms for the reduction of the number of points required to represent a digitised line or its caricature. The Canadian Cartographer, 10(2), 1973, pp. 112–123.
- [**Egenhofer et al., 1994**] M.J. Egenhofer, E. Clementini et P. Di Felice. Evaluating inconsistencies among multiple representations. In Proceedings of the 6th International Symposium on Spatial Data Handling (SDH'94), Volume 2, Edinburgh, Scotland, UK, 1994, pp. 901–920.
- [**Egenhofer et Herring, 1990**] M.J. Egenhofer et J.R. Herring. A mathematical framework for the definition of topological relationships. In Proceedings of the 4th International Symposium on Spatial Data Handling (SDH'90), Vol.2, Zurich, Switzerland, 1990, pp. 803–813.
- [**Egenhofer et Mark, 1995**] M.J. Egenhofer et D. Mark. Modelling conceptual neighbourhoods of topological line-region relations. International Journal of Geographical Information Science (IJGIS), 9(5), , Taylor and Francis, London, pp. 555–565.
- [**Ferber, 1995**] J. Ferber. Les systèmes multi-agents. Interéditons, 1995.
- [**Frank et Egenhofer, 1988**] M. J.Egenhofer et A. Frank. Object oriented database technology for GIS. In Proceedings of GIS/LIS'88, Annual Conference and Exposition, ASPRS (Ed.), Falls Church, VA, USA, 1988.
- [**Gabay et Doytsher, 1994**] Y. Gabay et Y. Doytsher. Automatic adjustment of line maps. In the proceedings of GIS/LIS'94, Annual Conference and Exposition. Phoenix, Arizona, USA, October 25-27, 1994, pp. 333–341.
- [**Gacogne, 1997**] L. Gacogne. Eléments de logique floue. Hermès, Paris, 1997, ISBN 2-86601-618-1.
- [**Gagnon, 1993**] P. Gagnon. Concepts fondamentaux de la gestion du temps dans les SIG. Mémoire de Maîtrise, Université Laval, Québec, Québec, Canada, 1993, 136 pages.
- [**GIS/Trans Ltd, 1995**] GIS/Trans Ltd. GIS Conflation and Its Automation. Technical Report GIS/Trans, Ltd, Cambridge, MA, 1995.
- [**Harvey et Vauglin, 1996**] F. Harvey et F. Vauglin. Geometric Match-processing : Applying Multiple Tolerances. In Proceedings of the 7th International Symposium on Spatial Data Handling (SDH'96), Delft, Netherlands, Kraak et Molenaar (Eds.), 1996, pp. 155–171.
- [**Hornsby et Egenhofer, 1997**] K. Hornsby et M. J. Egenhofer. Qualitative representation of change. In Proceedings of International Conference COSIT'97, Spatial Information Theory – A Theoretical Basis for GIS, S. C. Hirtle et A. U. Frank (Eds.), Springer, Laurel Highlands, Pennsylvania, October 1997, pp. 15–33.
- [**Hornsby et Egenhofer, 1998**] K. Hornsby et M. J. Egenhofer. Identity-based change operations for composite objects. In Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98), Vancouver, BC, Canada, edited by Poiker, T.K., and Chrisman, N., (Burnaby: International Geographical Union), 1998, pp. 88–97.

- [**Huet et al., 1996**] C. Huet, M. Tonon, H. Le Men, et O. Jamet. Towards an automatic updating process for land-cover cartography using SPOT images interpretation. In *The European Symposium on Satellite Remote Sensing III*, Taormina, 1996.
- [**IGN, 1997a**] IGN. Spécifications détaillées de la BD topographique. Version 3.1, Edition 4, IGN/SIT/Equipe produit BDTopo, Réf. SIT/97/0155, © IGN, Paris, Janvier 1997, 244 pages.
- [**IGN, 1997b**] IGN. Spécifications de la carte au 1 : 25 000 type 93. Edition 2, Révision 2, IGN/SIT/Projet BDTopo25, Réf. SIT/97/0877, © IGN, Paris, 30 mai 1997, 53 pages.
- [**IGN, 1998**] IGN. Descriptif de contenu Route500. Edition 2, © IGN, Paris, juin 1998, 21 pages.
- [**IGN, 1999**] IGN. Spécification de contenu BDCarto. Version 2, Rev. 4, IGN/SDT/Produit BDCARTO, Réf. SDT/99/-2322, © IGN, Paris, 30 juin 1999, 136 pages.
- [**IHO, 1996**] International Hydrographic Organisation. IHO transfer standard for digital hydrographic data - Publication S-57, Edition 3.0, November 1996, 126 pages.
- [**ISO, 1999**] ISO/TC 211. CD 15046-18 Geographic information – Part 18: Encoding. ISO/TC 211 N 709 Committee draft for comments, edited by ISO/TC 211 Secretariat, (Oslo: Norwegian Technology Standard Institution), 76 pages.
- [**Jamet, 1998**] O. Jamet. Comparaison Vecteurs-Images pour la détection des changements des bâtiments d'une base de données topographiques. Mémoire de thèse de doctorat de l'Ecole Nationale Supérieure des Télécommunications, spécialité Signal et Images, Paris, France, 11 décembre 1998, 204 pages.
- [**Jensen et Snodgrass, 1996**] C.S. Jensen et R.T. Snodgrass. Semantics of Time-varying Information. *Information Systems*, 21(4), 1996, pp. 311–352.
- [**Kemppäinen, 1994**] H. Kemppainen. Modelling updates propagation in spatial databases. In *International Archives of Photogrammetry and Remote Sensing*, 30 (4), 1994, pp. 625–633.
- [**Kilpeläinen, 1994**] T. Kilpeläinen. Updating multiple representation geodata bases by incremental generalization. In *International Archives of Photogrammetry and Remote Sensing*, 30 (3/1), 1994, pp. 440–447.
- [**Kilpeläinen, 1995**] T. Kilpeläinen. Updating Multiple Representation Geodata Bases By Incremental Generalization. In *Geo-Information-Systeme*, Wichmann, pp. 13–18.
- [**Kilpeläinen, 1997**] T. Kilpeläinen. Multiple representation and generalization of geo-databases for topographic maps. *Publications of the Finnish Geodetic Institute*, no 124, 1997, ISBN 951-711-211-4.
- [**Kuhn, 1994**] W. Kuhn. Defining semantics for spatial data transfers. In *Proceedings of the 6th International Symposium on Spatial Data Handling (SDH'94)*, Edinburgh, Scotland, UK, pp. 973–987.
- [**Langran, 1992**] G. Langran. *Time in Geographic Information Systems*. Taylor & Francis, London, 1992.
- [**Langran et Chrisman, 1988**] G. Langran et N. R. Chrisman. A framework for temporal geographic information. *Cartographica*, 25(3), 1988, pp. 1–14.
- [**Laurini, 1996**] R. Laurini. Raccordement géométrique de bases de données géographiques fédérées. *Ingénierie des systèmes d'informations*, vol.4, no. 3, 1996, pp. 361–388.
- [**Laurini et Thompson, 1992**] R. Laurini and D. Thompson. *Fundamentals of Spatial Information Systems*, (London: Academic Press), 1992.
- [**Lemarié, 1996**] C. Lemarié. Etat de l'art sur l'appariement. Rapport technique DT/9600022/S-RAP, IGN, Service de la Recherche, Saint Mandé, France, juillet 1996, 37 pages.
- [**Lemarié et Raynal, 1996**] C. Lemarié et L. Raynal. Geographic data matching: First investigations for a generic tool. In *Proceedings of GIS/LIS'96, Annual Conference and Exposition*. Denver, Colorado, USA, November 19-21, 1996, pp. 405–420.
- [**Lemarié et Bucaille, 1998**] C. Lemarié et O. Bucaille. Spécifications d'un module générique d'appariement de données géographiques. Actes du 11^{ème} congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA'98), Volume 2, Clermont-Ferrand, France, 20-22 janvier 1998, pp. 397–406.
- [**Lemon et Masters, 1997**] D. B. Lemon et E. Masters. The Nature and Management of Positional Relationships within Spatial Databases. In *Proceedings of the 2nd annual conference of Geocomputation'97*, University of Otago, New Zealand, August 26-29, 1997, pp. 15–23.

- [**Lemon, 1998**] D. B. Lemon. The nature and management of positional relationships within a local government geographic information system. Reports from School of Geomatic Engineering, University of New South Wales, Sidney, Australia, Ref. UNISURV S-53, 1998, 276 pages.
- [**Longley et al., 1999**] P. A. Longley, M. F. Goodchild, D. J. Maguire et D. W. Rhind (Eds.). Geographical Information Systems – Principles and Technical Issues. Volume 1, Second Edition, Jon Wiley & Sons, Inc., New York, 1999, ISBN 0471-33132-5.
- [**Lupien et Moreland, 1987**] A. E. Lupien et W. H. Moreland. A general approach to map conflation. In Proceedings of the ACSM/ASPRS Annual Convention & Exposition, Technical Papers, AutoCarto 8, Baltimore, Maryland, USA, March 29-April 3, 1987, pp. 630–639.
- [**Lynch et Salford, 1985**] M. Lynch et A. Salford. Conflation : Automated map compilation. A video game approach. In Proceedings of Auto Carto 7, Washington D.C, USA, 1985, pp. 343–352.
- [**Michalski et al., 1984**] R. Michalski, J. Carbonell et T. Mitchell. Machine Learning – An artificial intelligence approach, Springer-Verlag, Berlin, 1984.
- [**MRN, 1992**] Ministère de l'énergie et des ressources naturelles du Québec. Prototype sur l'intégration et l'échange de données à référence spatiale, rapport technique du Ministère de l'énergie et des ressources naturelles du Québec, Gouvernement du Québec, 1992.
- [**Network Working Group, 1992**] Network Working Group. The MD5 message-digest algorithm. RFC 1321, April 1992, edited by Rivest, R., (Cambridge: MIT). URL: <http://www.cis.ohio-state.edu/htbin/rfc/rfc1321.html> or <http://www.faqs.org/rfcs/rfc1321.html>.
- [**O₂, 1991**] The O₂ System. Communications of the ACM Society, vol. 34, num. 10.
- [**OGIS, 1999a**] OpenGIS Consortium. Web map server interface specification. OpenGIS project document 99-077r1, edited by Doyle, A., (Needham: International Interfaces inc.), 1999, 56 pages.
- [**OGIS, 1999b**] OpenGIS Consortium,. Geography Markup Language (GML). OGC RFC 11 13-Dec-1999, edited by Lake, R., (Vancouver: Galdos Systems Inc.), 1999, 37 pages.
- [**Parent et al., 1998**] C. Parent, S. Spaccapietra, E. Zimanyi, P. Donini, C. Plazanet, C. Vangenot, N. Rognon, et P. A. Crausaz. Mads, modèle conceptuel spatio-temporel. Revue Internationale de Géomatique, 7(3-4), Hermès, Paris, 1998, pp.:317–352.
- [**Poupart-Lavoie, 1997**] G. Poupart-Lavoie. Développement d'une méthode de transfert des mises à jour de données à référence spatiale. Mémoire de maîtrise (M. Sc.), Université Laval, Centre de Recherche en Géomatique, Québec, Canada, décembre 1997, 128 pages.
- [**Raynal et al., 1995**] L. Raynal, B. David et G. Schorter,. Building an OOGIS prototype: experiments with GeO2. In Proceedings of ACSM/ASPRS Annual Convention and Exposition (AutoCarto12), Charlotte, North Carolina, USA, Volume 4, 1995, pp. 137–146.
- [**Raynal, 1996**] L. Raynal. Some elements for modelling updates in topographic database. In Proceedings of GIS/LIS'96, Annual Conference and Exposition, Denver, Colorado, USA, November 19-21, 1996, pp. 1223–1232.
- [**Regnauld, 1997**] N. Regnauld. Généralisation du bâti : Structure spatiale de type graphe et représentation cartographique. Mémoire de thèse en informatique de l'université de Marseille, Marseille, France, 1997, 179 pages.
- [**Richard, 1993**] D. Richard. Documents de références FEIV 1.1. Rapport technique IGN, Saint Mandé, France, 1993.
- [**Rock, 1983**] I. Rock. The logic of perception. Cambridge, Massachusetts, MIT Press.
- [**Ruas, 1999**] A. Ruas. Modèle de généralisation de données géographiques à base de contraintes et d'autonomie. Mémoire de thèse de doctorat en Sciences de l'Information Géographique de l'Université de Marne-la-Vallée, France, 1999, 322 pages.
- [**Rumbaugh et al., 1991**] J. Rumbaugh, M. Blaha, W. Premerlani, F. Eddy, et W. Lorensen. Object-Oriented Modeling and Design. Prentice-Hall, Englewood Cliffs, 1991.
- [**Saalfeld, 1988**] A. Saalfeld. Conflation – Automated map compilation. International Journal of Geographical Information Systems (IJGIS), 2(3), pp. 217–228.
- [**Scholl et al., 1991**] M. H. Scholl, C. Laasch, M. Tresch. Updatable Views in Object-Oriented Databases. In Proceedings of the 2nd DOOD Conference, Munich, Germany, December 1991, 19 pages.

- [**Scholl et al., 1996**] M. Scholl, A. Voisard, J.P. Peloux, L. Raynal et P. Rigaux. SGBD Géographiques – Spécificités. International Thomson Publishing France, Paris, 1996, ISBN 2-84180-051-2.
- [**Snodgrass, 1992**] R. T. Snodgrass. Temporal Databases. Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Lecture Notes in Computer Science, Springer-Verlag, 639, pp. 22–64.
- [**Spéry, 1999**] L. Spéry. Historique et mise à jour de données géographiques: application au cadastre français. Mémoire de thèse de doctorat en Géographie de l'Université d'Avignon et des Pays de Vaucluse, 1999, 236 pages.
- [**Tellez et Servigne, 1997**] B. Tellez et S. Servigne. Updating Urban Database with Aerial Photographs: A Common Structuring Methodology. Computers, Environment and Urban Systems, vol. 21, no 2, mars 1997, pp. 133–145.
- [**Ubeda et Egenhofer, 1997**] T. Ubeda et M.J. Egenhofer. Topological Error Correcting in GIS. Advances in Spatial Databases (SSD'97), Berlin, Germany, Scholl et Voisard (Eds.), LNCS 1262, Springer-Verlag, 1997, pp. 283–297.
- [**Uitermark et al., 1998**] H. Uitermark, P. van Oosterom, N. Mars, M. Molenaar. Propagating updates: Finding Corresponding objects in a multi-source environment. In Proceedings of the 8th International Symposium on Spatial Data Handling (SDH'98), Vancouver, BC, Canada, edited by Poiker, T.K., and Chrisman, N., (Burnaby: International Geographical Union), 1998, pp. 580–591.
- [**van Oosterom, 1994**] P. van Oosterom. An R-tree based map-overlay algorithm. In Proceedings of the Fifth European Conference on Geographical Information Systems (ECGIS'94), Paris, France, March 1994, pp. 318–327.
- [**van Wijngaarden et al., 1997**] F. van Wijngaarden, J. van Putten, P. van Oosterom et H. Uitermark. Map Integration – Update Propagation in a Multi-Source Environment. In Proceedings of ACM GIS'97, Las Vegas, Nevada, USA, 1997, pp. 71–76.
- [**Vauglin et Bel Hadj Ali, 1998**] F. Vauglin, et A. Bel Hadj Ali. Geometric matching of polygonal surfaces in GIS. ASPRS-RTI Annual Conference, Tampa, Floride, USA, March 30-April 3, 1998, pp. 1511–1516.
- [**W3C, 1998**] World Wide Web Consortium. Extensible Markup Language (XML) 1.0. W3C Recommendation 10-February-1998. URL : <http://www.w3.org/TR/1998/REC-xml-19980210>.
- [**Worboys, 1994**] M.F. Worboys. Unifying the Spatial and Temporal Components of Geographical Information. In Proceedings of the 6th International Symposium on Spatial Data Handling, Edinburgh, Scotland, UK, 1994, pp. 505–517.
- [**Worboys et al., 1990**] M.F. Worboys, H.M. Hearnshaw et D.J. Maguire. Object-Oriented Data Modeling for Spatial Databases. International Journal of Geographical Information Systems, 4(4), pp. 369–383.

Publications

- [**Badard, 1998a**] T. Badard. Mise à jour générique des bases de données géographiques : Panorama des problèmes et premières investigations. Actes des 1ères journées Ré-ingénierie des Systèmes d'Information (RSI'98), Inforsid (Eds.), Lyon, France, 1-2 Avril, 1999, pp. 109–118.
- [**Badard, 1998b**] T. Badard. Premières pistes pour une mise à jour générique des bases de données géographiques. Bulletin d'information de l'IGN, n° 69 (1998/2), IGN, Paris, France, 1998, pp. 25–32.
- [**Badard, 1998c**] Extraction des mises à jour dans les BDG – De l'utilisation de méthodes d'appariement. Revue internationale de géomatique, vol. 8, n° 1-2, Hermès, Paris, 1998, pp. 121–147.
- [**Badard, 1998d**] T. Badard. Towards a generic updating tool for geographic databases. In Proceedings of GIS/LIS'98, Annual Conference and Exposition, AAG, ACSM, ASPRS, GITA, URISA (Eds.), Fort Worth, Texas, USA, November 1998, pp. 352–363.
- [**Badard, 1999**] T. Badard. On the automatic retrieval of updates in geographic databases based on geographic data matching tools. In Proceedings of the 19th International Cartographic Conference (Ottawa'99), ICA/ACI (Eds.), Ottawa, Ontario, Canada, 1999, pp. 47–56.
- [**Badard et Lemarié, 1999**] T. Badard et C. Lemarié. Propagating updates between geographic databases with different scales. Chapter 10 of Innovations in GIS VII: GeoComputation, Atkinson, P. and Martin, D. (Eds.), Taylor and Francis, London, 1999, 12 pages.
- [**Badard et Lemarié, 2000**] T. Badard et C. Lemarié. Propagating updates between geographic databases. In submission to the International Journal of Geographical Information Science (IJGIS), Taylor and Francis, London, 2000, 12 pages.
- [**Badard et Richard, 2000a**] T. Badard et D. Richard. Using XML for the exchange of updating information between geographical information systems. To appear in Computers, Environment and Urban Systems (CEUS), Elsevier Science, Amsterdam, 2000, 12 pages.
- [**Badard et Richard, 2000b**] T. Badard et D. Richard. Livraison des évolutions pour les BDG – Une proposition autour de XML. En soumission à la revue internationale de géomatique, Hermès, Paris, 2000.
- [**Badard et Spéry, 2000**] T. Badard et L. Spéry. Problèmes liés à l'échange et à l'intégration des mises à jour dans les bases de données géographiques. A paraître dans un volume de Information Géographique et Aménagement du Territoire, Traité de Géomatique, Hermès, Paris, 2000, 21 pages.