



Modélisation des contraintes d'intégrité dans les systèmes d'information environnementaux

François Pinet

► To cite this version:

François Pinet. Modélisation des contraintes d'intégrité dans les systèmes d'information environnementaux. Génie logiciel [cs.SE]. Habilitation à Diriger les Recherches de l'Université Blaise Pascal, Ecole Doctorale Sciences pour l'Ingénieur, Clermont-Ferrand (Spécialité Informatique), 2010. tel-02593346

HAL Id: tel-02593346

<https://hal.inrae.fr/tel-02593346>

Submitted on 15 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MEMOIRE

Habilitation à Diriger les Recherches

de l'Université Blaise Pascal, Ecole Doctorale Sciences pour l'Ingénieur, Clermont-Ferrand

(Spécialité Informatique) soutenue publiquement le 1^{er} avril 2010

par François Pinet

Chargé de Recherche, Cemagref

Modélisation des Contraintes d'Intégrité dans les Systèmes d'Information Environnementaux

Rapporteurs :

Ki-Joune Li Professeur - Université Nationale de Pusan

Michel Mainguenaud Professeur - INSA de Rouen

Michel Schneider Professeur - Université Blaise Pascal

Examineurs :

Jérôme Gensel Professeur - Université Pierre Mendès-France

Pierrick Givone Directeur de la Direction de la Stratégie et de la Recherche du Cemagref

Farouk Toumani Professeur - Université Blaise Pascal

Unité de Recherche TSCF - Technologie et Système d'information pour les agrosystèmes, Clermont-Ferrand
Thème de Recherche MOTIVE - MODélisation, Technologie de l'Information et Viabilité pour l'Environnement

Résumé

Le présent mémoire est composé de deux parties.

La première décrit l'essentiel de mon parcours professionnel, depuis la préparation de mon doctorat jusqu'à aujourd'hui. Cette partie résume mes activités scientifiques et indique les résultats marquants. Sont notamment précisés, les projets scientifiques et les collaborations initiés.

La seconde partie du mémoire présente les travaux de recherche menés dans le cadre des trois thèses de doctorat encadrées jusqu'à maintenant. Ces recherches portent sur la modélisation des contraintes d'intégrité dans les Systèmes d'Information (SI) environnementaux. Le contrôle des contraintes d'intégrité permet d'assurer la qualité des données d'un SI. Il est donc essentiel de disposer de méthodes pour représenter et contrôler ces contraintes dans les SI environnementaux. Une particularité des SI pour l'environnement est que les données que ces systèmes manipulent sont, par nature, géoréférencées. Dans ce cadre, le but principal de nos travaux était de fournir des moyens expressifs pour modéliser les contraintes d'intégrité des bases de données spatiales, ainsi que de mettre en place des outils pour les contrôler. Le challenge était de trouver un compromis entre le pouvoir d'expression offert par une méthode et sa complexité d'utilisation. Notre étude nous a amené à proposer des extensions spatiales d'OCL (Object Constraint Language). Nous décrivons dans ce mémoire, les solutions proposées pour les données spatiales certaines ou incertaines. Nous nous sommes concentrés sur les contraintes impliquant des relations topologiques dans leur formalisation. Nous nous sommes intéressés à la fois aux SI reposant sur des bases de données transactionnelles et aux systèmes décisionnels reposant sur des entrepôts de données. Plusieurs exemples dans le domaine de l'environnement sont présentés dans le mémoire.



Remerciements

Je remercie tous les membres du jury. Je remercie Pr. Jérôme GENSEL, Dr. Pierrick GIVONE, Pr. Ki-Joune LI, Pr. Michel MAINGUENAUD, Pr. Farouk TOUMANI. Je remercie tout spécialement le Pr. Michel SCHNEIDER pour son accueil depuis mon arrivée à Clermont-Ferrand en 2002 et pour sa collaboration scientifique.

Je remercie tous mes collègues du Cemagref pour leur collaboration, leur aide et leur soutien. Entre autres, un grand merci à Jean-Pierre CHANET, Frédéric VIGIER, André MIRALLES, Sandro BIMONTE, Catherine ROUSSEY, Stephan BERNARD, Gil DE SOUSA, Kamal BOULIL, Hadj MAHBOUBI, Lotfi BEJAOU, Magali DUBOISSET, Vincent SOULIGNAC, Vincent ABT, Daniel BOFFETY, Jean-Claude CHAMPOMIER, Géraldine ANDRE, Aurélien JACQUOT. Un grand merci aux collègues des services administratifs et notamment Christine ALAIN, Christiane ALBARET, Colette CADIOU, Eliane SIMON, Catherine ROULON. Je tiens à remercier Emmanuel HUGO, Directeur de l'UR TSCF, Guillaume DEFFUANT, animateur du TR MOTIVE, et Anne RIZAND, Directrice du centre du Cemagref de Clermont Ferrand.

Je remercie également, tous mes autres collègues, chercheurs français ou étrangers avec lesquels j'ai travaillé.

Enfin, un chaleureux merci à ma famille pour leur soutien de tous les jours : mes parents, ma sœur, mon frère, ma femme et mes enfants.

Sommaire

Partie 1 : Parcours professionnel

| | |
|------------------------------|----|
| A. Curriculum Vitae | 13 |
| B. Résumé des activités..... | 21 |
| C. Publications | 24 |

Partie 2 : Présentation des travaux de recherche

| | |
|--|----|
| Introduction | 33 |
| 1 Modélisation ER et OO des systèmes d'information spatiale | 37 |
| 1.1 Les besoins en formalismes adaptés à l'information spatiale | 37 |
| 1.2 Historique..... | 39 |
| 1.3 Synthèse des principales propositions existantes..... | 42 |
| 1.3.1 Représentation des objets spatiaux | 43 |
| 1.3.2 Modélisation des relations topologiques | 46 |
| 1.3.3 Modélisation de l'évolution temporelle des objets spatiaux..... | 47 |
| 1.4 Positionnement des recherches initiées..... | 48 |
| 1.5 Synthèse des fonctionnalités des différents formalismes..... | 50 |
| 2 Synthèse des méthodes opérationnelles pour spécifier les contraintes d'intégrité spatiale..... | 53 |
| 2.1 Introduction..... | 53 |
| 2.2 Principaux moyens existants pour modéliser les contraintes d'intégrité..... | 54 |
| 2.2.1 Outils pour la spécification de contraintes topologiques | 54 |
| 2.2.2 Formalismes et langages expressifs pour la spécification de contraintes..... | 57 |
| 2.3 Conclusion | 59 |
| 3 Modélisation de contraintes d'intégrité topologique impliquant des régions certaines..... | 61 |
| 3.1 Introduction..... | 61 |
| 3.2 Extensions spatiales d'OCL..... | 61 |
| 3.3 Implémentation | 65 |

| | | |
|-------|--|-----|
| 3.3.1 | Présentation d'OCL2SQL..... | 66 |
| 3.3.2 | Principe de l'extension spatiale d'OCL2SQL | 66 |
| 3.4 | Expérimentation sur des données environnementales..... | 67 |
| 3.4.1 | Présentation du contexte | 67 |
| 3.4.2 | Exemple de contraintes..... | 69 |
| 3.5 | Conclusion et pistes de recherche à développer..... | 72 |
| 4 | Modélisation de contraintes d'intégrité topologique impliquant des régions vagues..... | 77 |
| 4.1 | Introduction..... | 77 |
| 4.2 | Les régions incertaines et leurs relations topologiques..... | 77 |
| 4.2.1 | Modèles antérieurs à nos propositions..... | 77 |
| 4.2.2 | Un nouveau modèle de régions incertaines | 79 |
| 4.3 | Extension d'OCL pour les régions spatiales incertaines..... | 81 |
| 4.4 | Expérimentation sur des données environnementales..... | 83 |
| 4.5 | Conclusion et pistes de recherche à développer..... | 84 |
| 5 | Contraintes d'intégrité dans les entrepôts de données..... | 87 |
| 5.1 | Introduction..... | 87 |
| 5.2 | Présentation de l'étude de cas | 91 |
| 5.3 | Exemples de contraintes | 94 |
| 5.4 | Conclusion et pistes de recherche à développer..... | 98 |
| | Conclusion et perspectives générales | 101 |
| | Références | 103 |

Liste des figures

| | |
|---|----|
| Figure 1. Résumé des différentes thèses de doctorat..... | 35 |
| Figure 2. Un modèle représenté en UML et décrivant des données spatiales | 38 |
| Figure 3. Un modèle représenté avec le formalisme proposé par (Lbath, 1997) (OMEGA) et décrivant les mêmes données spatiales que le modèle de la Figure 2 | 39 |
| Figure 4. Deux exemples de transformation d'un modèle Perceptory en UML..... | 44 |
| Figure 5. Exemple de diagramme CONGOO..... | 46 |
| Figure 6. Exemple de pictogramme temporel du formalisme Perceptory (figure issue de (Bédard et al., 2004)) | 47 |
| Figure 7. Exemple de relations temporelles avec MADS et GeoOOA | 48 |
| Figure 8. Interface de l'outil proposé par (Servigne et al., 2000; Ubeda & Servigne, 1997)..... | 56 |
| Figure 9. Interface de l'outil de (Cockcroft, 1998, 2001; Cockcroft, 2004) | 56 |
| Figure 10. Les entreprises et leurs employés..... | 58 |
| Figure 11. huit relations topologiques possibles entre deux régions simples (Egenhofer & Herring, 1992) (Egenhofer & Franzosa, 1991)..... | 62 |
| Figure 12. Les bâtiments de mairie et leur commune | 63 |
| Figure 13. Les centres ville et leurs îlots de bâtiments..... | 64 |
| Figure 14. Exemple d'un centre ville et de ses îlots de bâtiments..... | 65 |
| Figure 15. Entrées / sorties du générateur OCL2SQL spatial | 67 |
| Figure 16. Diagramme de classes de SIGEMO | 69 |
| Figure 17. Exemple de parcelles d'épandage se trouvant sur des départements limitrophes du département d'évaluation du dossier (63) | 71 |
| Figure 18. Couplage d'OCL2SQL avec les outils de type interface visuelle de définition de contraintes spatiales | 73 |
| Figure 19. Une région spatiale vague selon le modèle du jaune d'œuf..... | 78 |
| Figure 20. Deux régions spatiales floues (exemple issu de (Schneider, 2008))..... | 78 |
| Figure 21. Une région spatiale vague à la frontière partiellement vague | 79 |
| Figure 22. Cinq exemples de relations entre deux régions vagues selon le modèle <i>QMM</i> de (Bejaoui et al., 2008a)..... | 80 |
| Figure 23. Correction de la géométrie des parcelles | 83 |

| | |
|---|-----|
| Figure 24. Intégration de trois sources de données stockant chacune le même objet | 86 |
| Figure 25. Exemple d'entrepôt de données (issu de (Trujillo et al., 2001)) | 88 |
| Figure 26. Une contrainte de disjonction (Pinet & Schneider, 2009a) entre deux niveaux d'une dimension | 90 |
| Figure 27. Etapes principales de la conception des entrepôts | 90 |
| Figure 28. Modèle multidimensionnel | 91 |
| Figure 29. Une analyse présentant les résultats des zones d'épandage par producteur, commune et catégorie de produit ; la fonction d'agrégation utilisée est l'union spatiale | 92 |
| Figure 30. Deux nouvelles classes | 95 |
| Figure 31. Ajout d'une dimension temporelle..... | 97 |
| Figure 32. Exemple de chaîne de production d'entrepôt de données..... | 102 |

Liste des tableaux

| | |
|--|----|
| Tableau 1. Exemple des combinaisons de pictogrammes possibles dans le formalisme Perceptory (tableau issu de (Bédard et al., 2004))..... | 45 |
| Tableau 2. Buts atteints par les différents formalismes proposés..... | 51 |
| Tableau 3. Positionnement de notre proposition par rapport à l'existant..... | 73 |
| Tableau 4. Exemple de ligne incertaine selon le modèle <i>QMM</i> (Bejaoui et al., 2009) | 84 |
| Tableau 5. Les différentes analyses spatialisées possibles à partir du modèle de la Figure 28..... | 93 |

Partie 1 : Parcours professionnel

A. Curriculum Vitae

François PINET

né le 3 mai 1973 à Saint-Chamond

Chargé de Recherche 1^{ère} classe, affecté au centre de Clermont-Ferrand du Cemagref (site d'Aubière)

Domaines de compétences disciplinaires : informatique, géomatique.

Derniers diplômes obtenus

2002 : Thèse de doctorat en informatique (INSA de Lyon).

1997 : DEA Informatique de Lyon (Université Lyon 1),
Magistère Informatique et Modélisation (ENS Lyon).

Fonctions

- Depuis Sept. 2003 : Chargé de recherche au Cemagref, centre de Clermont-Ferrand, au sein de l'équipe SI (Systèmes d'Information), UR TSCF¹ et Thème de Recherche MOTIVE². 1^{ère} classe depuis Sept. 2007.
- Sept. 2002 - Août 2003 : Attaché Temporaire d'Enseignement et de Recherche (mi-temps) à l'Université Blaise Pascal (Clermont-Ferrand), rattaché au LIMOS³, équipe "Bases de Données et Systèmes d'Information".
- Fév. 1999 - Août 2002 : Thèse de doctorat en informatique au sein du laboratoire LISI⁴ (Equipe "Système d'Information Géographique" - INSA de Lyon) et de la société CIRIL SA, Villeurbanne ; contrat CIFRE. Les travaux portaient sur une méthodologie de conception d'applications géomatiques communicantes et une extension JAVA de l'outil de génie logiciel AIGLE commercialisé par la filiale Business Geographic de CIRIL SA (<http://www.business-geografic.com>).
- Nov. 1998 - Janv. 1999 : CDD à l'INSERM (Hôpital Louis Pradel - CHU de Lyon) dans l'équipe de recherche du Professeur Rubel. Les travaux portaient sur la conception d'outils informatiques pour la visualisation de données médicales.
- Nov. 1997 - Août 1998 : Service National (Base aérienne de Limonest, Rhône).

¹ UR TSCF : Unité de Recherche « Technologie et Systèmes d'information pour les agrosystèmes » de Clermont Fd

² MOTIVE : MOdélisation, Technologie de l'Information et Viabilité pour l'Environnement

³ LIMOS : Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes, Clermont Université

⁴ LISI : Laboratoire d'Ingénierie des Systèmes d'Information, INSA Lyon

Activités d'enseignement

Toute mon activité d'enseignement concerne l'informatique.

a) Formation initiale

- Depuis Sept. 2003, entre 50 et 70 heures annuelles d'enseignement réalisées en Auvergne :
 - o "Systèmes d'Information Géographique" (cours) en Master 2 "Stratégies Internet et Pilotage de Projets" (Vichy),
 - o "Bases de données avancées et Internet" (cours/TP de 2003 à 2008) en Master 2 bioinformatique, puis en Master 2 Analyse et modélisation des données pour les sciences du vivant (Clermont-Ferrand),
 - o "Bases de données" (cours/TD/TP) à l'école d'ingénieur ISIMA (de 2007 à 2009),
 - o "Sciences pour l'ingénieur au service des agrosystèmes" (cours) à l'Ecole Doctorale SPI,
 - o "Entrepôts de données" (cours) aux étudiants de l'ENITA de Bordeaux (effectués dans les locaux du Cemagref de Clermont Fd).
- Année universitaire 2002-2003 : ATER à l'Université Blaise Pascal, 96h d'enseignement, en licence informatique (C++, systèmes), en DESS bioinformatique (bases de données), ainsi qu'à l'école d'ingénieur CUST (Java).
- Entre 1999 et 2002, en moyenne 60 heures de vacations par an, principalement à : l'Ecole d'ingénieur CPE (Lyon), l'IUT de Bourg-en-Bresse (Université Lyon 1), et en Deug (Université Lyon 2).

b) Formation continue

- Définition et mise en place de la formation continue « UML pour la modélisation de systèmes environnementaux » avec André Miralles (UMR TETIS⁵ - Montpellier), Gilles Bonnet (Direction Scientifique du Cemagref) et Petraq Papajorgji (Univ. Floride - USA). Il s'agit d'une formation annuelle d'une semaine donnée depuis 2007 à la Maison de la Télédétection de Montpellier. Les cours ont été suivis principalement par des ingénieurs et chercheurs du Cemagref, de l'Engref, de l'INRA, du Cirad, du CNRS et du Ministère de l'Agriculture.
- Entre 2004 et 2007, formateur « Conception de bases de données » à l'INSERM (formation interne annuelle d'une semaine dispensée à Kremlin Bicêtre).

Animations scientifiques

- Cofondateur et animateur du réseau international de recherche « Design of Environmental & Agricultural Systems » avec André Miralles et Petraq Papajorgji. Le réseau a été mis en place en 2006. Il vise à réunir les chercheurs de différents instituts intéressés à collaborer sur le thème de la conception des systèmes d'information environnementaux. Aujourd'hui, il est composé d'une trentaine de chercheurs de 11 instituts (5 pays) : Agricultural University of Athens (Grèce), Cemagref, Cirad, Embrapa (Brésil), CAO⁶ de l'Univ. de Floride, CRG⁷ (Canada), ITK, LIMOS, LIRIS⁸, Univ. Federal de Viçosa (Brésil), Université du Québec à

⁵ UMR TETIS : Unité de Mixte de Recherche « Territoire, Environnement, Télédétection, Information Spatiale »

⁶ CAO : Center of Applied Optimization

⁷ CRG : Centre de Recherche en Géomatique

⁸ LIRIS : Laboratoire d'InfoRmatique en Image et Systèmes d'information

Montréal (Canada). Quatre ateliers ont été organisés dans le cadre du réseau. Site Web : <http://deas.research.free.fr>

Comités scientifiques et éditoriaux

- Membre du comité de rédaction de la « Revue Internationale de Géomatique » (revue créée en 1991 - Hermès/Lavoisier - France). geo.e-revues.com
- Cofondateur et corédacteur en chef avec Petraq Papajorgji du « International Journal of Agricultural & Environmental Information Systems » (IGI Global / Idea Group Inc - USA). Le « Editorial Review Board » est actuellement composé de 45 chercheurs venant de différents pays. www.igi-global.com/IJAEIS
- Membre du conseil scientifique de l' « International Conference on Discovering, Mining and Managing Complex Agricultural Data 2010 » (Crete)
- Membre du comité scientifique de l' « International Workshop on Geographical Analysis, Urban Modeling, Spatial Statistics », en lien avec l' « International Conference on Computational Science and its Applications » (ICCSA) en 2010 (Fukuoka - Japon)
- Membre du comité scientifique des 6èmes Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA) en 2010 (Djerba)
- Membre du comité scientifique de la conférence internationale « Free and Open Source Software for Geospatial » (FOSS4G) en 2009 (Sydney)
- Membre du comité scientifique de la « 5th International Conference on Signal-Image Technology & Internet-Based Systems » (SITIS - Track Web-Based Information Technologies & Distributed Systems) en 2009 (Marrakech)
- Membre du comité scientifique du « Quatrième Atelier sur les Systèmes Décisionnels » (ASD) en 2009 (Jijel - Algérie)
- Membre du comité scientifique du « Colloque International de Géomatique et d'Analyse Spatiale » (SAGEO), depuis 2007
- Membre du comité scientifique de la Conférence Québeco-Française pour le développement de la géomatique (CQFD-Géo) en 2007 (Clermont-Ferrand)
- Membre de liaison à l'International Commission of Agricultural Engineering depuis 2006 (Section VII – Information Systems). L'objectif de la commission internationale fondée en 1930 est de stimuler et promouvoir le développement des sciences et technologies pour l'agriculture. Les actions de la commission passent, par exemple, par la diffusion d'un e-journal et par l'organisation de conférences internationales. Dans ce contexte, je me suis impliqué au niveau du comité scientifique de l'International Conference of Agricultural Engineering (Brésil) en 2008 - Section « interdisciplinary issues », organisé par la commission www.acquacon.com.br/icae/en/

Depuis 2003, j'ai été sollicité pour reviewer une cinquantaine d'articles scientifiques pour différents journaux et colloques (nationaux et internationaux) : International Journal of Geographical Information Science (Taylor&Francis), Revue "Computers & Geosciences" (Elsevier), Revue

"Knowledge-Based Systems" (Elsevier), Revue "Operational Research" (Springer), Journal of Location Based Services (Taylor&Francis), Revue "Optimisation, Modeling & Software" (Taylor&Francis), International Journal of Agricultural and Environmental Information Systems (IGI Global), Revue Internationale de Géomatique (Hermès/Lavoisier), Revue des Nouvelles Technologies de l'Information (Cépaduès), Encyclopedia of Geoinformatics (Idea Group), Congrès INFORSID 2003 et 2005 (France), International Symposium on Web et Wireless Geographic Information 2004 à 2006 (Corée du Sud, Suisse, Hong Kong), Colloque COSI 2005 (Algérie), International Conference on Advanced Information Systems Engineering CAISE 2007 (Norvège), Conférence Québéco-Française pour le Développement de la Géomatique 2007 (France), SAGEO 2007 à 2010 (France), Conférence BDA 2007 (France), International Conference of Agricultural Engineering 2008 (Brésil), ASD 2009 (Algérie), FOSS4G 2009 (Australie), EDA 2010 (Djerba).

Comités d'organisation de colloques et d'ateliers

- Organisateur prévu des 7^{ème} journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA). Organisation planifiée au Cemagref de Clermont-Ferrand en 2011
- Organisateur avec Sandro Bimonte et André Miralles de l'atelier « Systèmes d'Information et de Décision pour l'Environnement » (SIDE) associé au 27ème Congrès INFORSID, Toulouse, 26 mai 2009
- Organisateur avec André Miralles de l'atelier « Systèmes d'Information et de Décision pour l'Environnement » (SIDE) associé au 26ème Congrès INFORSID, Fontainebleau, 27 mai 2008
- Membre du comité d'organisation du « Colloque International de Géomatique et d'Analyse Spatiale » (Sageo 2007), Engref, Clermont-Ferrand, 18-19 Juin 2007
- Organisateur avec Gilles Bonnet du Workshop « AgriUML », 15-17 Mai 2006, Cemagref, Clermont-Ferrand
- Organisateur avec Petraq Papajorgji du Workshop « Modeling Agricultural Systems using an MDA Approach » (MAS), 11 novembre 2006, en collaboration avec la « 5th International Conference AFITA », Bangalore, India
- Membre du comité d'organisation des 4èmes journées d'Extraction et de Gestion des Connaissances, Clermont-Ferrand, Université Blaise Pascal, 20-23 Janvier 2004
- Membre du comité d'organisation du « First International Workshop on TeleGeoprocessing » (1999), INSA Lyon, 6-7 Mai, 1999

Principales missions collectives et d'intérêt général au sein du Cemagref

- Mise en place, dépouillement et synthèse d'une enquête nationale concernant l'organisation des bases de données au sein du Cemagref. Cette enquête interne a été réalisée auprès des 21 Unités de Recherche et de 6 UMR de l'établissement entre 2005 et 2006. Travaux en collaboration avec Gilles Bonnet (Direction Scientifique du Cemagref) et Nicolas Moyroud (Cemagref Montpellier).

- Membre élu à la Commission Administrative Paritaire (CAP) des Chargés de Recherche de l'établissement (2004-2008). Les CAP traitent des questions relatives aux carrières individuelles du personnel. Les CAP sont chargées principalement d'examiner des situations individuelles, mais elles peuvent aussi émettre des motions à caractère collectif. Les modalités générales de traitement des carrières des agents y sont par exemple discutées.

Projets financés (depuis Sept. 2003)

Voici les projets dans lesquels j'ai eu un niveau d'implication important, que cela soit dans la définition des problématiques scientifiques à aborder ou dans les solutions apportées.

- 2009-2011 : Disp'Eau (8^{ème} appel du Fonds Unique Interministériel) : Développement d'un système d'aide à la décision pour l'irrigation de la vigne.
 - o Partenaires : Société ITK, Association Climatologique de l'Hérault, Cemagref, Chambre d'Agriculture de l'Hérault, CIRAD, Geocarta, INRA, Les Vignobles Foncalieu, Netafim France, Vignerons du Sieur d'Arques.
 - o L'objectif du projet est de développer un système d'aide à la décision permettant de piloter l'irrigation pour la vigne. Notre rôle dans le projet est de mettre en place l'acquisition des données via un réseau sans fil de capteurs et leur centralisation dans un Système d'Information.
 - o Le projet est porté par Valérie Auffray et Eric Jallas (Société ITK). 3 permanents de l'équipe SI de l'UR TSCF sont impliquées dans le projet (Jean-Pierre Chanet, Gil De Sousa, François Pinet). Je suis le correspondant du projet au niveau de l'équipe.
- 2008-2011 : SIE Pesticide (Programme national "Pesticide" - Appel à projets du Cemagref) : Systèmes d'Information et Entrepôts de données pour l'analyse des résidus de pesticides.
 - o Partenaires : UMR TETIS, UR TSCF, UR ADBX⁹, UR HHLY¹⁰ (Cemagref - Montpellier, Clermont-Ferrand, Bordeaux, Lyon).
 - o Le but premier du projet est de concevoir un entrepôt de données spatiales et temporelles, capitalisant les informations sur les pesticides et leurs utilisations en agriculture.
 - o Le projet est porté par André Miralles (UMR TETIS). 5 permanents de l'équipe SI de l'UR TSCF sont impliquées dans le projet (Stephan Bernard, Sandro Bimonte, Jean-Pierre Chanet, Gil De Sousa, François Pinet). Je suis responsable du pilotage scientifique du projet au niveau de l'équipe.
- 2007-2009 : OTAG (Projet européen – 6ème PCRD) : Operational management and geodecisional prototype to track and Trace AGricultural production.
 - o Partenaires : UR TSCF, UMR TETIS, Embrapa (Brésil), Cirad (Montpellier), GAAP¹¹ (Univ. Laval – Québec).
 - o L'objectif du projet était d'obtenir une traçabilité fine des mouvements de la production animale agricole. De nouvelles technologies de l'information et de la communication ont été développées dans le projet pour enregistrer les déplacements

⁹ UR ADBX : Unité de Recherche « Aménités et Dynamiques des espaces ruraux » du Cemagref de Bordeaux

¹⁰ UR HHLY : Unité de Recherche « Hydrologie – Hydraulique » du Cemagref de Lyon

¹¹ GAAP : laboratoire de Géomatique Agricole et APpliquée

- de nombreux troupeaux de bovins. Le développement d'une large base de données de traçabilité a été étudié en lien avec un système d'aide à la décision. Des expérimentations ont été réalisées au Brésil.
- Le projet était porté par Gérard Chuzel (Direction du Cemagref) et Dominique Didelot (Chargé de mission - Cemagref). 4 permanents de l'équipe SI ont été impliqués (Stephan Bernard, Sandro Bimonte, Jean-Pierre Chanet, François Pinet). J'ai travaillé plus particulièrement sur la définition du Système d'Information et d'aide à la décision. Avec Sandro Bimonte, chargé de recherche nouvellement recruté dans l'équipe, nous sommes en train de définir un sujet de thèse dans la continuité du thème du projet.
 - Site Web : <http://www.otag-project.org>
- 2006-2009 : FORUM (Projet ANR): A Semantic Based Mediation System for Large Scale Applications.
 - Partenaires : LIRMM¹² (Montpellier), LIMOS, UR TSCF, IRISA¹³ (Lannion), LIRIS (Lyon).
 - L'objectif était de définir un système de médiation pour interroger un grand nombre de sources de données ; l'expérimentation a été faite sur environ 300 sources dans les domaines environnementaux et agricoles.
 - Le projet était porté par Zohra Bellahsene (LIRMM). Au sein de l'équipe SI, 2 permanents (Stephan Bernard, François Pinet) travaillaient sur le projet. J'étais responsable du WP « application agricole » du projet, ainsi que du pilotage scientifique au niveau de l'équipe.
 - 2005 : EnviroBus (Contrat industriel) : Capteurs pour la mesure urbaine de NO2 et cartographie des résultats.
 - Partenaires : Syndicat Mixte des Transports en Commun de Clermont-Ferrand, LASMEA¹⁴, LIMOS, UR TSCF
 - L'objectif a été d'utiliser des capteurs de NO2 réalisés au LASMEA, pour acquérir des mesures urbaines géoréférencées. Pour cela, les capteurs ont été embarqués dans des bus (sur Clermont-Ferrand et Aubière), et une cartographie des mesures a été réalisée.
 - 2 permanents (Jean-Pierre Chanet, François Pinet) ont participé au projet dans l'équipe SI. Mon rôle dans le projet a été de concevoir la cartographie des mesures.
 - 2004-2008 : Développement d'un volet scientifique autour du projet d'ingénierie SIGEMO - projet de SI pour la Gestion des épandages de Matières Organiques à l'échelle nationale, conçu par le Cemagref (UR TSCF) en lien avec le Ministère de l'Agriculture, le Ministère de l'Environnement, l'ADEME, les Agences de l'Eau.
 - Le projet d'ingénierie SIGEMO est porté par Vincent Soullignac dans l'équipe SI du Cemagref. J'ai développé entre 2004 et 2008, un projet de recherche en lien avec SIGEMO. Ce projet scientifique impliquant deux doctorants de l'équipe (Magali Duboisset et Lotfi Bejaoui) consistait à fournir des moyens de contrôle de contraintes d'intégrité pour la base de données spatiales de SIGEMO. Le projet SIGEMO se poursuit toujours en lien avec l'UR TSCF.

¹² LIRMM : Laboratoire d'Informatique, de Robotique et de Micro-électronique de Montpellier

¹³ IRISA : Institut de recherche en informatique et systèmes aléatoires

¹⁴ LASMEA : Laboratoire des Sciences et Matériaux pour l'Electronique et d'Automatique

- 2004-2006 : STAR (Programme d'Action Intégré): Innovative Spatial Model and Structure (SPIN).
 - o Partenaires : IRENAV¹⁵, STEM lab¹⁶ (Corée du Sud), UR TSCF, LIMOS
 - o Il s'agit d'un projet sur les structures et modèles spatiaux innovants. L'équipe SI du Cemagref a contribué à la définition des contraintes d'intégrité du modèle. Le porteur du projet était Eric Saux (IRENAV). J'étais responsable du pilotage scientifique du projet au niveau de l'équipe SI.
- 2003-2004 : RAHA (Programme GEOMATIC du CNRS): Réseau Ad-Hoc Agricole et bases de données contextuelles.
 - o Partenaires : LIMOS, UR TSCF, IGN
 - o Le projet concernait les réseaux Ad-Hoc de communication sans fil en milieu rural et les bases de données contextuelles associées. Mon travail portait sur la création du modèle de bases de données spatio-temporelles.

Expertises

Deux expertises sont actuellement en cours de mise en place et sont prévues pour l'année 2010 :

- Participation à une expertise pour l'INRA. L'étude portera sur l'évaluation d'un logiciel développé par l'INRA. L'expertise sera réalisée en lien avec Myoung-Ah KANG, Maître de Conférences au LIMOS. Durée de l'expertise : 2 semaines de scientifique.
- Expertise prévue pour une société Clermontoise. L'étude porte sur les entrepôts de données spatiaux. Durée de l'expertise : 5 semaines de scientifique. Participants : François Pinet, Sandro Bimonte (équipe SI de l'UR TSCF)

Activités d'encadrement d'étudiants

Etudiants de doctorat

- décembre 2004 - novembre 2007 : Magali Duboisset (ingénieur de l'Institut Supérieur d'Informatique, de Modélisation et de leurs Applications de Clermont-Ferrand).
 - o Thèse de doctorat « Un système de contraintes d'intégrité OCL pour les bases de données spatiales : application à un système d'information pour l'épandage agricole », soutenue le 3 décembre 2007.
 - o Encadrement : François Pinet, Michel Schneider (LIMOS).
 - o Inscription à l'ED SPI de l'Univ Blaise Pascal (UBP).
 - o Laboratoire d'accueil : UR TSCF (Cemagref).
 - o Financement : 50% bourse Cemagref + 50% bourse Région Auvergne.
- janvier 2006 - mai 2009 : Lotfi Bejaoui (titulaire du DEA Informatique de Paris Dauphine).
 - o Thèse de doctorat « Qualitative topological relationships for objects with possibly vague shapes: implications on the specification of topological integrity constraints in

¹⁵ IRENAV : Institut de Recherche de l'Ecole NAVale de Brest

¹⁶ STEM lab : Spatio-TEMporal database Lab

transactional spatial databases and in spatial data warehouses », *soutenue le 25 mai 2009*.

- Encadrement : François Pinet, Michel Schneider (LIMOS), Yvan Bédard (Centre de Recherche en Géomatique - CRG, Université Laval, Québec).
 - Inscription à l'ED SPI et à l'Université Laval ; convention de cotutelle signée entre l'UBP, l'Université Laval et le Cemagref.
 - Laboratoires d'accueil : UR TSCF (Cemagref) et CRG.
 - Financement : 50% bourse Cemagref + 50% CRG.
- depuis octobre 2009 : Kamal Boulil (titulaire d'un Master Informatique de l'Université de Dijon).
- Thèse de doctorat « Conception d'entrepôts de données environnementales ».
 - Encadrement : François Pinet, Sandro Bimonte (UR TSCF), Michel Schneider (LIMOS).
 - Inscription à l'ED SPI.
 - Laboratoires d'accueil : UR TSCF (Cemagref).
 - Financement : 50% bourse Cemagref + 50% bourse Région Auvergne.
- depuis mars 2010 : Guillaume Barbier (diplômé de l'Institut National d'Agronomie Paris Grignon).
- Thèse de doctorat « Vers une architecture générique de modélisation mécaniste des cultures ».
 - Encadrement : David Hill (LIMOS), François Pinet.
 - Inscription à l'ED SPI.
 - Laboratoire d'accueil : LIMOS.
 - Financement : convention CIFRE avec la société ITK de Montpellier.

Etudiants de niveau Bac+5

Depuis 2000 :

- Encadrement d'étudiants de DEA et Masters recherche
 - 1 étudiant du DEA DISIC de Lyon, 2 étudiants du DEA Informatique de l'UBP, 1 étudiant du Master Recherche Informatique de l'UBP.
- Encadrement d'étudiants d'écoles d'ingénieurs et de DESS
 - 2 étudiants de l'ISIMA, 1 étudiant du DESS SIAD de l'UBP.

B. Résumé des activités

Parcours scientifique

Ma thèse en informatique (1999-2002) réalisée à la fois dans le milieu académique et le milieu industriel (thèse CIFRE), portait sur le développement d'un outil de génie logiciel (AIGLE JAVA) pour la conception de Systèmes d'Information Géographique (SIG). Cette solution commerciale a principalement permis de faciliter la production d'applications géomatiques communicantes (Web et mobiles). Fort de cette formation, je poursuivis en 2002, mes recherches au sein du LIMOS de Clermont-Ferrand, accueilli en tant qu'ATER dans l'équipe « Base de Données » dirigée par Michel Schneider. Au cours de cette année d'ATER, j'ai, entre autres, travaillé avec le Cemagref sur un projet lié à la modélisation de bases de données géographiques pour les réseaux de communication sans fil en milieu agricole. En septembre 2003, je fus recruté au Cemagref sur un poste de chargé de recherche (centre de Clermont-Ferrand). J'étais le premier docteur dans une équipe de recherche naissante très dynamique dirigée à l'époque par Frédéric Vigier. Cette équipe Système d'Information (équipe SI) était composée d'ingénieurs et de doctorants recevant au niveau local un appui scientifique du LIMOS. Plusieurs ingénieurs permanents de l'équipe préparaient des thèses de doctorat. L'équipe était impliquée dans de nombreux projets et avait engagé des recherches sur deux thèmes complémentaires : la conception de SI agri-environnementaux et l'acquisition de données par des réseaux sans fil de capteurs pour alimenter ces SI. Une caractéristique importante des SI ciblés par ces thèmes est qu'ils traitent des données géoréférencées.

A mon entrée au Cemagref, ma mission première a été de développer un sujet de recherche s'intégrant aux thèmes existants. J'ai donc étudié les différents thèmes et projets déjà présents dans l'équipe. Cette étude m'a mené à faire monter en puissance les deux sujets scientifiques suivants, avec l'équipe et ses partenaires, tout spécialement Michel Schneider : *le contrôle de la qualité des SI par des contraintes d'intégrité et la conception d'entrepôts de données dans le domaine agri-environnemental*. Dans les systèmes agri-environnementaux, la prise en compte du caractère spatial des données est essentielle. Cette prise en compte nécessitait de nouvelles recherches scientifiques. Pour traiter les sujets scientifiques ciblés, nous avons développé de nouvelles approches de génie logiciel basées sur UML (Unified Modeling Language). Un volet de notre activité a aussi porté sur la conception de SI agri-environnementaux communicants ; j'ai participé à différents projets portant sur ce thème. A l'heure d'aujourd'hui, l'équipe est composée d'une quinzaine de personnes dont quatre docteurs permanents.

Principaux résultats

A l'heure d'aujourd'hui, mes principaux résultats sont les suivants :

- Le renforcement des compétences de l'équipe SI dans le domaine des SI agri-environnementaux avec plusieurs projets en appui.
- 2 thèses de doctorat encadrées et soutenues, 2 thèses en cours.
- Les liens tissés avec de nouveaux partenaires et la constitution d'une communauté internationale autour de la conception de systèmes agri-environnementaux ; cette communauté regroupe plusieurs chercheurs de laboratoires de différents pays. Plusieurs publications scientifiques communes ont été réalisées.
- La structuration de cette communauté par la mise en place et l'animation, avec Petraq Papajorgji (Univ. Floride) et André Miralles (UMR TETIS) du réseau international de recherche « Design of Environmental & Agricultural Systems » regroupant une trentaine de chercheurs de 11 instituts.
- Une dissémination des connaissances des nouvelles techniques de génie logiciel pour les SI agri-environnementaux, par la mise en place de plusieurs Workshops nationaux et internationaux, et d'une formation spécifique (organisée avec Gilles Bonnet de la Direction Scientifique du Cemagref, André Miralles et Petraq Papajorgji).

- La création de la revue « International Journal of Agricultural & Environmental Information Systems » (IGI Journal) dont je suis corédacteur en chef avec Petraq Papajorgji.

Recherche partenariale et projets

Après une étude théorique des moyens à mettre en œuvre et l'évaluation des solutions possibles, l'élaboration de techniques basées sur OCL/UML¹⁷ est ressortie comme la solution la plus pertinente pour modéliser les contraintes d'intégrité dans les SI environnementaux. Ces contraintes permettent de contrôler la qualité des données. Les données environnementales étant spatialisées, il a fallu étendre OCL pour permettre de spécifier des contraintes sur des objets spatiaux, et intégrer l'utilisation d'un tel langage aux méthodes de conception pour les SI spatiaux (travaux de la thèse de Magali Duboisset). Je me suis aussi rapproché de l'Université Technologique de Dresde (Allemagne) qui travaille sur des outils de modélisation de contraintes d'intégrité avec OCL. Notre extension spatiale d'OCL a été implémentée dans l'outil de l'Université de Dresde. Cet outil a été mis en œuvre avec Vincent Soullignac (Cemagref Clermont-Ferrand) sur le Système d'Information pour les épandages appelé SIGEMO. Après une étude détaillée du contexte de l'épandage, une implantation des contraintes d'intégrité a été réalisée avec l'outil de l'Université de Dresde étendu au spatial. Nous avons donc une première version d'un outil pour évaluer la qualité des SI agri-environnementaux. En plus de l'application sur le projet SIGEMO, le projet STAR (2004-2006) a offert un autre champ d'application et a permis d'expérimenter l'outil développé sur d'autres types d'objets spatiaux environnementaux. Des travaux en lien avec la modélisation de SI associée aux réseaux sans fil ont aussi été effectués.

Certains des objets considérés dans les SI agri-environnementaux présentent un degré de complexité supplémentaire du fait de leurs limites incertaines ou de leur forte variabilité temporelle (un peuplement forestier, une zone humide,...). Il nous a semblé important d'aborder ce problème. Pour cela, avec la collaboration de la Direction Scientifique du Cemagref, je me suis rapproché de l'équipe d'Yvan Bédard du CRG qui était déjà un partenaire identifié de notre équipe. Nous avons alors initié la thèse de Lotfi Bejaoui, avec le CRG et le LIMOS. La collaboration avec le CRG nous a permis d'avancer plus rapidement sur les sujets scientifiques abordés, l'équipe québécoise ayant une bonne expérience des problèmes de qualité dans les bases de données spatiales.

En dehors de la modélisation des contraintes d'intégrité, j'ai aussi souhaité utiliser les propositions faites d'extension d'OCL pour décrire les propriétés des données spatiales de Systèmes d'Aide à la Décision agricole. Afin de pouvoir considérer un cas suffisamment complet en agriculture, je me suis rapproché de chercheurs du Joint Research Center de la Commission Européenne (Ispra en Italie). Nous avons expérimenté l'approche sur un projet de Système d'Aide à la Décision pour la Politique Agricole Commune. Nos résultats communs ont été publiés dans un journal international en 2006.

Afin d'être en mesure de répondre à de nouveaux besoins d'analyses des données, nous avons, plus récemment, choisi de fortement orienter une partie de nos activités sur les entrepôts de données en collaboration avec l'Université Laval. Sandro Bimonte, un nouveau chargé de recherche spécialiste de ce thème a été recruté dans l'équipe en février 2009. Nous collaborons très étroitement avec André Miralles (Cemagref Montpellier) sur ce sujet. Le premier projet financé dans le domaine est le projet SIE Pesticide. Une partie du projet européen OTAG traite aussi de ces aspects.

Avec différents chercheurs intéressés par nos sujets ou ayant participé à nos projets, nous avons commencé à former une communauté autour de la qualité des SI, et plus largement autour de la conception des systèmes agri-environnementaux. En 2006, nous avons décidé avec Petraq Papajorgji et André Miralles de « formaliser » cette communauté en mettant en place et en animant un réseau de

¹⁷ OCL : Object Constraint Language ; UML : Unified Modeling Language

recherche « Design of Environmental & Agricultural Systems » regroupant aujourd'hui une trentaine de chercheurs de 11 instituts de différents pays.

J'ai aussi travaillé sur des projets portant sur les SI agri-environnementaux communicants, notamment en lien avec les chercheurs de l'équipe travaillant sur les réseaux sans fil. J'ai poursuivi le projet RAHA au sein de l'équipe ; puis le projet EnviroBUS avec le LIMOS, le LASMEA et le Syndicat Mixte des Transports en Commun de Clermont-Ferrand. Je me suis chargé de la partie "bases de données" de ce projet. Cette partie consistait à réaliser des analyses et une cartographie de mesures environnementales géoréférencées de NO₂ relevées automatiquement par des systèmes embarqués. Le projet ANR FORUM a été monté et démarré avec nos partenaires en 2006. Le projet portait sur les toutes dernières technologies en matière d'interopérabilité de SI, et de leur adaptation possible à la sphère agricole. En plus de notre équipe et du LIMOS, le projet impliquait différentes équipes expertes en base de données, de plusieurs laboratoires d'informatique. Au-delà de la synergie scientifique, ceci nous permet d'obtenir auprès de la communauté scientifique nationale des SI, une bonne lisibilité dans les domaines des SI pour l'agriculture et l'environnement. Nous avons aussi établi une collaboration avec la société ITK dirigée par Eric Jallas, chercheur au Cirad. Avec ITK, nous commencerons une thèse CIFRE et démarrons le projet Disp'Eau (2009-2012). Dans ce projet, notre champ d'intervention porte sur la mise en place d'un réseau de capteurs et sur son couplage avec un Système d'Aide à la Décision pour l'irrigation de la vigne. Par ailleurs, pour renforcer nos activités dans le domaine de l'interopérabilité des SI et les ontologies, Catherine Roussey, Maître de Conférences à l'Université Lyon 1, s'est intégrée à l'équipe depuis septembre 2009 (en délégation).

C. Publications

| Type de publications | Nombre |
|---|---|
| <i>Revue à comité de lecture</i> | 13 (12 revues internationales et 1 revue nationale) |
| <i>Lecture Notes in Computer Science (Springer)</i> | 6 |
| <i>Chapitre d'ouvrage</i> | 4 |
| <i>Edition d'actes et d'ouvrage</i> | 3 |
| <i>Publication dans des actes de colloque</i> | 31 (24 internationaux et 7 nationaux) |
| <i>Critique d'ouvrage</i> | 1 |
| <i>Mémoire</i> | 2 |
| <i>Rapport</i> | 5 |

Revues à comité de lecture

1. BIMONTE, S., TCHOUNIKINE, A., MIQUEL, M., **PINET, F.** - 2010. When Spatial Analysis Meets OLAP: Multidimensional Model and Operators. To appear in: *International Journal of DataWarehousing and Mining*, IGI Global, 36 pages
2. BIMONTE, S., SCHNEIDER, M., MAHBOUBI, H., **PINET, F.**, CHANET, J.P. - 2010. Merging Spatial Data Cubes Using the GIS Overlay Operator. To appear in: *Journal of Decision Systems*, Lavoisier, 27 pages
3. BEJAOU, L., **PINET, F.**, BEDARD, Y., SCHNEIDER, M. - 2010. OCL for formal modelling of topological constraints involving regions with broad boundaries. *Geoinformatica* vol.14(3), Springer, p. 353 - 378
4. MIRALLES, A., **PINET, F.**, BEDARD, Y. - 2010. Object-oriented modelling of spatio-temporal environmental phenomena: an overview of today's needs and solutions. To appear in: *International Journal of Agricultural and Environmental Information Systems*, IGI Global, 18 pages
5. **PINET, F.**, SCHNEIDER, M. - 2010. Precise design of environmental data warehouses. To appear in: *Operational Research*, Springer, 24 pages
6. MAHBOUBI, H., BIMONTE, S., FAURE, T., DEFFUANT, G., CHANET, J.P., **PINET, F.** - 2010. Data warehouse and OLAP for Environmental Simulation Data. To appear in: *International Journal of Agricultural and Environmental Systems*, IGI Global, 22 pages
7. PAPAJOJGI, P., **PINET, F.**, MIRALLES, A., JALLAS, E., PARDALOS, P. - 2010. Modeling: a central activity for flexible information systems development in agriculture

and environment. *International Journal of Agricultural and Environmental Information Systems*, vol. 1(1), IGI Global, p. 1 - 25

8. **PINET, F., SCHNEIDER, M.** - 2009. A unified object constraint model for designing and implementing multidimensional systems. *Journal of Data Semantics*, vol. 13, Springer, p. 37 - 71
9. **BEJAOU, L., PINET, F., BEDARD, Y., SCHNEIDER, M.** - 2009. Qualified topological relations between spatial objects with possibly vague shape. *International Journal of Geographical Information Science*, Taylor & Francis, vol. 23(7), p. 877 - 921
10. **PINET, F., DUBOISSET, M., SCHNEIDER, M.** - 2009. Modélisation de contraintes d'intégrité spatiales avec OCL. *Revue Internationale de Géomatique*, vol.19(1), Lavoisier, p. 93 - 122
11. **PINET, F., DUBOISSET, M., SOULIGNAC, V.** - 2007. Using UML and OCL to maintain the consistency of spatial data in environmental information systems. *Environmental modelling & software*, vol. 22(8), Elsevier, p. 1217 - 1220
12. **HASENOHR, P., PINET, F.** - 2006. Modeling of a spatial DSS template in support to the Common Agricultural Policy. *Journal of Decision Systems*, vol. 15(2), Lavoisier, p. 181 - 196
13. **PINET, F., LBATH, A., VIGIER, F., SCHNEIDER, M.** - 2003. Conception de systèmes d'information à référence spatiale (SIRS) communicants pour les agrosystèmes. *Ingénieries – Eau, Agriculture, Territoires*, n° spécial Technologies pour les agrosystèmes durables, p. 143 - 155

Lecture Notes in Computer Science (Springer)

14. **PINET, F., SCHNEIDER, M.** - 2008. A semantic language for querying anonymous web sources. *Lecture Notes in Computer Science*, vol. 5178 (KES 2008), p. 106 - 116
15. **BEJAOU, L., PINET, F., SCHNEIDER, M., BEDARD, Y.** - 2008. An adverbial approach for the formal specification of topological constraints involving regions with broad boundaries. *Lecture Notes in Computer Science*, vol. 5231 (ER 2008), p. 383 - 396
16. **DUBOISSET, M., PINET, F., KANG, M.A., SCHNEIDER, M.** - 2007. A general framework to implement topological relations on composite regions. *Lecture notes in computer science*, vol. 4653 (DEXA 2007), p. 823 - 833
17. **DUBOISSET, M., PINET, F., KANG, M.A., SCHNEIDER, M.** - 2005. Precise modeling and verification of topological integrity constraints in spatial databases: from an expressive power study to code generation principles. *Lecture notes in computer science*, vol. 3716 (ER 2005), p. 465 - 482
18. **PINET, F., KANG, M.A., VIGIER, F.** - 2005. Spatial constraint modelling with a GIS extension of UML and OCL: application to agricultural information systems. *Lecture notes in computer science*, vol. 3511 (Metainformatics 2004), p. 160 - 178
19. **PINET, F., LBATH A.** - 2001. Semantics of Stereotypes for Type Specification in UML: Theory and Practice. *Lecture notes in computer science*, vol. 2224 (ER 2001), p. 339 - 353

Chapitres d'ouvrage

20. **VISOLI, M., BIMONTE, S., TERNES, S., CHANET, J.P., PINET, F.** - 2010. Towards a New Architecture for Spatial Decision Support System for Animals Traceability. To

appear in: *Computational Methods Applied to Agricultural Research: Advances and Applications*, IGI Global, 22 pages

21. ROUSSEY, C., **PINET, F.**, KANG, M.A., CORCHO, O. - 2009. How ontologies are used for software interoperability. Chapter to appear in: *Use of Ontologies to Support Information Interoperability*, Springer, 52 pages
22. **PINET, F.**, DUBOISSET, M., DEMUTH, B., SCHNEIDER, M., SOULIGNAC, V., BARNABE, F. - 2009. Constraints modeling in agricultural databases. Chapter in: *Advances in Modeling Agricultural Systems*, Springer, p. 55 - 66
23. **PINET, F.**, ROUSSEY, C., BRUN, T., VIGIER, F. - 2009. The use of UML as a tool for the formalisation of standards and the design of ontologies in agriculture. Chapter in: *Advances in Modeling Agricultural Systems*, Springer, p. 131 - 147

Editions d'actes et d'ouvrages

24. BIMONTE, S. (ed.), MIRALLES, A. (ed.), **PINET, F.** (ed.) - 2009. Actes de l'atelier "Systèmes d'Information et de Décision pour l'Environnement" (SIDE 2009), associé au 27ème Congrès INFORSID, Toulouse, 26 mai 2009. 105 pages
25. SCHNEIDER, M. (ed.), BATTON HUBERT, M. (ed.), JOLIVEAU, T. (ed.), LARDON, S. (ed.), **PINET, F.** (ed.) - 2008. Extraction et gestion de connaissances dans un contexte spatio-temporel : applications à la géomatique pour la gestion des territoires. Cépaduès éditions, Toulouse, 218 p.
26. **PINET, F.** (ed.), MIRALLES, A. (ed.) - 2008. Actes de l'atelier "Systèmes d'Information et de Décision pour l'Environnement" (SIDE 2008), associé au 26ème Congrès INFORSID, Fontainebleau, 27 mai 2008. 84 pages

Publications dans des actes de colloque

27. VISOLI, M., TERNES, S., **PINET, F.**, CHANET, J.P., MIRALLES, A., BERNARD, S., DE SOUSA, G. - 2009. Computational architecture of OTAG project. European Conference for Information Technologies in Agriculture, EFITA 2009, 6-8 July 09, Wageningen, NLD. 8 p.
28. BEJAOU, L., **PINET, F.**, SCHNEIDER, M., BEDARD, Y. - 2009. Qualitative topological relationships for objects with possibly vague shapes: implications on the specification of topological integrity constraints, International Symposium on Spatial Data Quality (ISSDQ-2009), July 5th-8th, St-John's Newfoundland, Canada
29. BERTOLOTTO, M., BIMONTE, S., DI MARTINO, S., FERRUCCI, F., LEANO, V., **PINET, F.** - 2009. Vers l'Analyse Spatio-Multidimensionnelle à l'aide du geobrowser Google Earth. Actes du 27ème Congrès INFORSID, Toulouse, 26 mai 2009
30. KARA, M., HOU, K.M., CHANET, J.P., ZHOU, H.Y., KANG, M.A., **PINET, F.** - 2007. Low cost differential GPS receivers (LCD-GPS): a local cooperative differential GPS solution. 7ème Conférence Internationale sur les NOuvelles TEchnologies de la REpartition (NOTERE'07), Marrakech, MAR, 4-8 juin, 2007. 6 pages
31. DEVILLERS, R., BEDARD, Y., GERVAIS, M., JEANSOULIN, R., **PINET, F.**, SCHNEIDER, M., BEJAOU, L., LEVESQUE, M.A., SALEHI, M., ZARGAR, A. - 2007. How to improve geospatial data usability: from metadata to quality-aware GIS community. AGILE Pre-Conference working group on spatial data usability, Aalborg, Denmark, DNK, 8 May 2007. 6 pages

32. DUBOISSET, M., **PINET, F.**, SCHNEIDER, M. - 2007. P-OCLcbm: a language to specify topological constraints on composite regions: definition and comparison with TRCR. 1st International Conference on Research Challenges in Information Science, Ouarzazate, MAR, 23-26 avril 2007. p. 373 - 378
33. BEJAOU, L., BEDARD, Y., **PINET, F.**, SAHELI, M., SCHNEIDER, M. - 2007. Logical consistency for vague spatiotemporal objects and relations. 5th International symposium on spatial data quality (ISSDQ 2007), Enschede, NLD, 13-15 June 2007. 8 pages
34. KARA, M., HOU, K.M., CHANET, J.P., ZHOU, H.Y., KANG, M.A., **PINET, F.** - 2006. Low cost differential GPS receivers (LCD-GPS): urban vehicle tracking. 5th IEEE International Information and Telecommunication Technologies Symposium (I2TS), Cuiabá , MT , Brazil , December 06-08, 2006 . 6 pages
35. **PINET, F.**, VENTADOUR, P., BRUN, T., PAPAJOJGI, P., ROUSSEY, C., VIGIER, F. - 2006. Using UML for ontology construction: a case study in agriculture. 7th Agricultural Ontology Service (AOS) Workshop on Ontology-Based Knowledge Discovery: Using Metadata and Ontologies for Improving Access to Agricultural Information (in collaboration with AFITA 2006), Indian Institute of Science, Bangalore , India , 9-10 November 2006, p. 735 - 739
36. DE SOUSA, G., CHANET, J.P., AMAMRA, A., HAO, J., KANG, M.A., **PINET, F.**, HOU, K.M. - 2005. Protocole de communication sans fil dédié : Obstacle Location-Aided Routing. Journée Véhicule Intelligent de l'Ecole Doctorale Sciences pour l'Ingénieur de Clermont-Ferrand, 27 juin 2005. 12 pages
37. HOU, K.M., CHANET, J.P., **PINET, F.**, VIGIER, F., AUFRERE, R., ZHOU, H., DE SOUSA, G., AMAMRA, A., SCHNEIDER, M., KANG, A. M. - 2005. For an Inter Vehicle Communication System Intelligent: CIVIC. Colloque Société de l'information bilan du programme interdisciplinaire de recherche du CNRS 2001 - 2005, 19-20-21 mai 2005, Lyon. 9 pages
38. DUBOISSET, M., **PINET, F.**, KANG, M.A., SCHNEIDER, M. - 2005. Integrating the calculus-based method into OCL: study of expressiveness and code generation. 2nd IEEE International DEXA Workshop on Geographic Information Management (GIM 2005), Copenhagen , DNK, 22-26 August 2005, pages 502 – 506
39. SOULIGNAC, V., GIBOLD, F., **PINET, F.**, VIGIER, F. - 2005. Spreading matter management in France within SIGEMO. 5th European Conference for Information Technologies in Agriculture (EFITA 2005), Vila Real, Portugal , 25-28 July 2005. 8 pages
40. HASENOHR, P., **PINET, F.**, KAY, S., GRENIER, G., SEVILA, F. - 2005. A case-study based methodology for modeling of spatial DSS using Spatial OCL . Joint Workshop on Decision Support Systems, Experimental Economics & e-Participation (GRAZ 2005), AUT, June 2005. 2 pages
41. VIGIER, F., CHANET, J.P., **PINET, F.**, BOFFETY, D. - 2005. Mobilisation et gestion des informations liées aux pratiques agricoles : systèmes d'information à référence spatiale partagés et moyens de communication interopérables. *Comptes-rendus de l'Académie d'Agriculture de France*, vol. 91, n° 3 (séance du 18 mai 2005), 8 pages
42. **PINET, F.**, VIGIER, F., LBATH, A. - 2004. Un formalisme adapté aux géomaticiens pour décrire la distribution et les échanges de données au sein des systèmes télégéomatiques. 5ème séminaire REGLIS Mise en place de systèmes d'information à références spatiales (SIRS), Montpellier, 13-14 novembre 2003. p. 127 - 139
43. CHANET, J.P., JAUDOIN, H., DUBOISSET, M., **PINET, F.** - 2004. Information systems technologies for sustainable management: a global approach. AGENG'2004 Information Systems Technologies for Sustainable Management: a Global Approach, Leuven , BEL, 11-15 September 2004. 8 pages

44. KANG, M.A., **PINET, F.**, SCHNEIDER, M., CHANET, J.P., VIGIER, F. - 2004. How to design geographic database? Specific UML profile and spatial OCL applied to wireless Ad Hoc networks. 7th Conference on Geographic Information Science (AGILE'2004), Heraklion, GRC, April 29-May 1 2004. p. 289 - 299
45. **PINET, F.**, KANG, M.A., JAUDOIN, H. - 2004. Link between the whole et its parts in UML representations of spatial aggregations: an application in the context of geographic databases. 7th Conference on Geographic Information Science (AGILE'2004) , Heraklion, GRC, April 29-May 1 2004. p. 523 - 532
46. SOULIGNAC, V., **PINET, F.**, GIBOLD, F. - 2004. Note sur SIGEMO : Système Informatisé de Gestion des Epanchages de Matières organiques. Congrès CASSINI 2004, Grenoble, 2-4 juin 2004. 2 pages
47. CHANET, J.P., BOFFETY, D., LI, J., **PINET, F.**, GERBE, P., SCHNEIDER, M., VIGIER, F., HOU, K.M. - 2003. Ad Hoc network for agriculture. 4th EFITA Conference, Debrecen-Budapest, HUN, 5-9 July 2003. 1 page
48. **PINET, F.**, LBATH, A. - 2003. Object-Oriented Modelling of Communicating GIS-based Applications. 6th AGILE European Conference on Geographic Information Science, Lyon, 24-26 avril 2003. p. 743 – 750
49. **PINET, F.**, LBATH, A. - 2001. An Algorithm for Stereotype Deduction in UML-Based Formalism and its Application in Geographic Information Systems. In: Proceedings of IEEE Symposium on Visual/Multimedia Approaches to Programming and Software Engineering (Human Centric Computing'2001), Stresa, Italy , September 2001. p. 339-353
50. **PINET, F.**, LBATH, A. - 2001. Des Applications Télégéomatiques Multi-Architectures pour Internet/Intranet/Extranet. Actes de CASSINI 2001 « Géomatique et Espace Rural », Montpellier, France, Septembre 2001. p. 287-288
51. LBATH, A., **PINET, F.** - 2000. The Development and Customization of GIS-Based Applications and Web-Based GIS Applications with the CASE Tool AIGLE. In: Proceedings of the 8th ACM Symposium on Advances in Geographic Information Systems (ACM GIS'2000), Washington D.C , USA , November 2000. p. 194-196
52. **PINET, F.**, LBATH, A. - 2000. A Visual Modelling Language for Distributed Geographic Information Systems. In: Proceedings of the 16th IEEE Symposium on Visual Languages (VL'2000), Seattle , Washington , USA , September 2000. p. 75-76
53. LBATH, A., **PINET, F.** - 2000. Automatic Generation of Geographic Urban Applications for the Web with the CASE Tool AIGLE. In: Proceedings of the 22nd Symposium on Urban and Regional Data Management (UDMS'2000), Netherlands , September 2000. Theme II, p.35-39
54. LBATH, A., **PINET, F.** - 2000. Towards Conceptual Modelling of TeleGeoProcessing Applications. In: Proceedings of the 2nd Symposium on TeleGeoProcessing (TeleGeo'2000), Nice, France , May 2000. p. 25-39
55. **PINET, F.**, LBATH, A. - 2000. An Extension of the CASE Tool AIGLE for Web-Based GIS Applications. In: Proceedings of the 2nd Symposium on TeleGeoProcessing (TeleGeo'2000), Nice, France , May 2000. p. 41-51
56. **PINET, F.**, LBATH, A. - 2000. AIGLE: a CASE Tool for Web-Based GIS Applications. In: Proceedings of the 1st Workshop on Emerging Technologies for Geo-Based Applications (ASCONA'2000), Ascona, Switzerland, May 2000. p. 77-101
57. BOULICAUT, J.F., MARCEL, P., **PINET, F.**, RIGOTTI, C. - 1998. Spreadsheet Generation from Rule-Based Specifications. In: Proceedings of the 6th International Workshop on Deductive Database and Logic Programming (DDL' 98), Manchester , UK , June 1998. p. 59-70

Critique d'ouvrage

58. **PINET, F.** - 2007. Book review of "Software Engineering Techniques Applied to Agricultural Systems: an Object-Oriented and UML Approach". *Optimization Methods and Software*, vol. 22(2), Taylor&Francis, p. 361-363

Mémoires

59. **PINET, F.** - 2002. Vers une Méthode de Conception pour la TéléGéomatique : Intégration et Validation au sein d'un Atelier de Développement. Mémoire de thèse de doctorat, le 7 octobre 2002, LISI INSA Lyon. 254 pages
60. **PINET, F.** - 1997. Analyse Statique d'un Langage de Requêtes sous Contraintes. Mémoire de stage de DEA, juillet 1997, LISI INSA Lyon. 31 pages

Rapports

61. **PINET, F.,** BERNARD, S. - 2009. Rapport d'expérimentation : projet FORUM. 16 pages (projet FORUM)
62. VISOLI, M., TERNES, S., MIRALLES, A., CHANET, J.P., **PINET, F.,** DE SOUSA, G., PIRES, P.P., AMARAL, T. - 2008. UML Model of the OTAG local system. 8 pages (projet OTAG)
63. VISOLI, M., TERNES, S., CHANET, J.P., **PINET, F.,** DE SOUSA, G., MIRALLES, A. - 2008. OTAG information system. Projet Européen OTAG (Operational Management and Geodecisional Prototype to Track and Trace Agricultural Production) Specific Support Action du 6ème PCDR. 15 pages (projet OTAG)
64. **PINET, F.,** BERNARD, S., BARNABE, F. - 2007. Cas d'utilisation : projet FORUM. 41 pages (projet FORUM)
65. HOU, K.M., CHANET, J.P., **PINET, F.,** VIGIER, F., LI, J.J., AUFRERE, R., ZHOU, H., DE SOUSA, G., AMAMRA, A., SCHNEIDER, M., KANG, M.A. - 2004. Vers une Communication Inter Véhicule Intelligente Coopérative : CIVIC. 10 pages (projet RAHA)

Partie 2 : Présentation des travaux de recherche

Introduction

Recherche sur les contraintes d'intégrité

Le présent mémoire fait le tour des principales contributions scientifiques que nous avons apportées dans le domaine de la qualité des SI agri-environnementaux, notamment dans le cadre des thèses encadrées jusqu'à maintenant. La partie scientifique du présent mémoire se focalise en effet sur les résultats de recherche attendus et obtenus dans ces thèses dans le domaine des contraintes d'intégrité.

Le contrôle des contraintes d'intégrité permet d'assurer la qualité des données d'un SI. Les données qui ne vérifient pas ces contraintes sont considérées comme erronées. Il est donc important de disposer de méthodes pour modéliser et contrôler ces contraintes dans les SI agri-environnementaux. Une caractéristique importante de ce type de SI est que les données que ces systèmes manipulent sont, par nature, géoréférencées. L'objectif des travaux présentés ici était de fournir des moyens expressifs pour spécifier les contraintes d'intégrité des bases de données spatiales, ainsi que de mettre en place des outils pour les contrôler. Le challenge est de trouver un compromis entre le pouvoir d'expression offert par une méthode et sa complexité d'utilisation. Par le passé, plusieurs propositions pour décrire les contraintes d'intégrité des bases de données spatiales ont été faites. Ces méthodes fournissent d'excellents résultats pour l'expression de contraintes d'un niveau de complexité moyen mais il était nécessaire d'investiguer des solutions plus expressives pour modéliser la complexité de certaines contraintes spatiales.

Notre étude nous a amené à utiliser et étendre OCL (Object Constraint Language) afin de permettre l'expression de contraintes spatiales. OCL est un langage pour l'expression de contraintes appliquées à des diagrammes UML (Unified Modeling Language). Il fait maintenant partie intégrante du standard UML. Un nombre grandissant de concepteurs de SI utilise ce langage en complément des diagrammes de classes d'UML pour exprimer les invariants des modèles par exemple. Le langage OCL peut être facilement intégré à n'importe quel formalisme Orienté-Objet ou Entité-Relation utilisé pour la conception de SI spatiaux. Ces avantages nous ont donc amené à baser notre proposition sur OCL. Nous avons produit une solution logicielle pour générer, à partir d'expressions OCL, des mécanismes de contrôle de contraintes d'intégrité dans les bases de données spatiales. Dans ce mémoire, nous présentons l'utilisation de cette solution pour vérifier les données d'un SI pour l'épandage agricole.

Dans ce mémoire, nous décrivons principalement les solutions proposées pour spécifier des contraintes sur des données spatiales certaines ou incertaines. Il s'agit de types d'objets courants dans le domaine agri-environnemental. Nous nous sommes concentrés sur les contraintes impliquant des relations topologiques dans leur formalisation. Nous nous sommes intéressés à la fois aux SI reposant sur des bases de données transactionnelles et aux systèmes décisionnels reposant sur des entrepôts de données.

Les avancées scientifiques décrites dans ce mémoire sont :

- *Une représentation innovante (basée sur OCL) des contraintes d'intégrité des SI et des entrepôts de données spatiales.* Ces contraintes d'intégrité permettent de décrire les relations topologiques qui doivent être respectées entre les objets spatiaux stockés. La modélisation des contraintes impliquant des objets spatiaux certains et incertains est considérée.
- *L'implémentation d'un outil logiciel permettant de vérifier que les contraintes d'intégrité modélisées sont respectées dans les SI.* L'expérimentation de cet outil sur un SI agri-environnemental est présentée.
- *La proposition d'un nouveau modèle de représentation d'objets spatiaux incertains expressif et facile à manipuler.*

- *la proposition d'utiliser OCL pour modéliser des contraintes complexes (y compris spatiales) dans les entrepôts de données.*

Travaux des doctorants

La présente partie scientifique de ce mémoire se concentre sur les recherches menées dans le cadre des thèses de doctorat encadrées (passées et en cours). Mes différents encadrements d'étudiants de doctorat sont les suivants.

La thèse de Magali Duboisset a été initiée en décembre 2004 et soutenue le 3 décembre 2007. Le sujet portait sur la conception de contraintes d'intégrité pour contrôler la qualité des données spatiales. Le résultat de ces recherches a été utilisé sur un SI national pour le suivi des épandages.

La thèse de Lotfi Bejaoui fut initiée en janvier 2006 en cotutelle avec l'Université Laval et l'Université Blaise Pascal. Elle fut soutenue le 25 mai 2009. Elle se focalisait plus spécifiquement sur les contraintes d'intégrité portant sur des objets spatiaux incertains.

La thèse de Kamal Boulil a été lancée en octobre 2009 sur les formalismes pour la conception d'entrepôt de données spatio-temporelles. L'application de la thèse est un entrepôt de données environnementales dans le cadre d'un projet de conception de SI sur les pesticides.

Ces trois premières thèses ont été accueillies dans notre équipe du Cemagref.

Avec David Hill, Professeur au LIMOS et la société ITK (Montpellier), nous avons aussi mis en place une thèse CIFRE. L'objectif est la mise en œuvre d'un outil de génie logiciel basé sur UML pour faciliter la production de systèmes de simulation en agronomie. Cette thèse réalisée par Guillaume Barbier, ingénieur diplômé de l'Institut National d'Agronomie Paris Grignon, a été acceptée par l'ANRT et a commencé en mars 2010.

La thèse de Marcos Visoli, ingénieur permanent à l'institut de recherche agricole du Brésil (Embrapa), est en cours de mise en place. Marcos Visoli est venu en France suivre un Master Recherche à l'Université Blaise Pascal en 2008-2009 et a travaillé avec nous sur le projet Européen OTAG. J'ai suivi son stage de Master réalisé dans l'équipe. Son doctorat concernera la conception d'entrepôts de données spatiales pour la traçabilité de la production agricole bovine. L'objectif est d'améliorer les outils actuels de traçabilité, qui sont amenés à stocker des quantités de données de plus en plus grandes. Nous discutons avec l'Embrapa de la date de début de la thèse, qui se déroulera en France, dans l'équipe (fin 2010 ou début 2011).

La Figure 1 résume les différentes thèses de doctorat passées, présentes et en cours de démarrage.

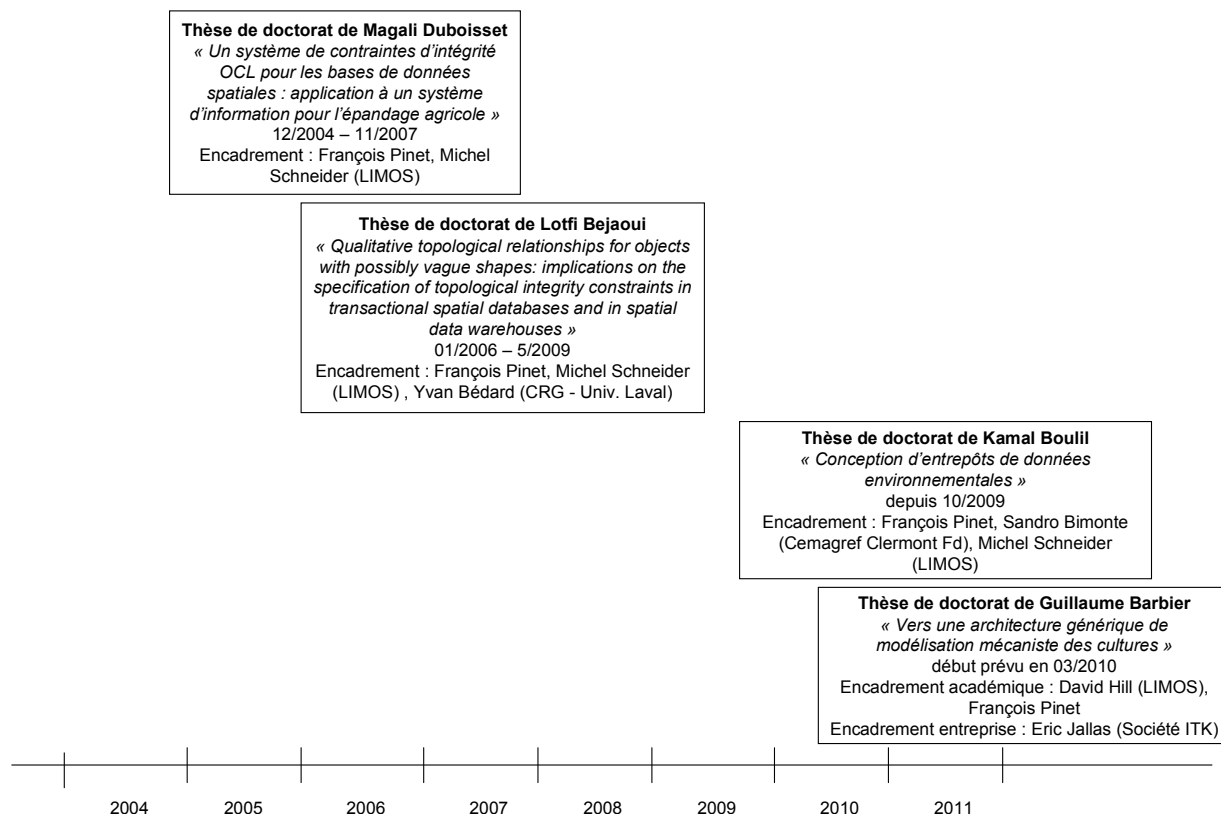


Figure 1. Résumé des différentes thèses de doctorat

Plan de la présentation

Lorsque je suis arrivé dans l'équipe, j'ai étudié les formalismes de conception pour les SI spatiaux, des recherches étant déjà initiées sur ce thème dans l'équipe (notamment en lien avec la thèse de doctorat préparée par André Miralles). Le Chapitre 1 du présent mémoire résume cette étude et décrit comment elle m'a amené à positionner mes recherches sur la qualité des SI agri-environnementaux, et plus spécifiquement sur la représentation et le contrôle des contraintes d'intégrité spatiale.

Le Chapitre 2 dresse une synthèse des solutions opérationnelles existantes pour, d'une part, spécifier et contrôler les contraintes spatiales, et d'autre part, modéliser des contraintes complexes dans les bases de données non-spatiales. Il ressort qu'avant nos propositions, il n'existait pas de solution dédiée pour représenter et contrôler des contraintes d'intégrité spatiale complexes.

Le Chapitre 3 décrit ensuite nos premières contributions sur la modélisation des contraintes d'intégrité spatiale avec OCL (thèse de Magali Duboisset). Nous présentons l'outil logiciel que nous avons produit et l'expérimentation que nous avons réalisée dans le domaine agri-environnemental.

Le Chapitre 4 présente notre contribution concernant la représentation des contraintes d'intégrité pour les bases de données d'objets spatiaux incertains (thèse de Lotfi Bejaoui). Nous décrivons tout d'abord notre modèle d'objets incertains en le positionnant par rapport aux modèles existants. L'implémentation et l'expérimentation de notre approche sont ensuite résumées.

Le Chapitre 5 est plus prospectif. Il explique l'approche que nous envisageons pour modéliser les contraintes d'intégrité (y compris spatiale) avec OCL dans les entrepôts de données (thèse en cours de Kamal Bouil).

Une partie « conclusion et perspectives générales » clôture le mémoire. Cette partie résume les points marquants des contributions scientifiques en indiquant les pistes futures qui nous semblent les plus importantes.

1 Modélisation ER et OO des systèmes d'information spatiale

1.1 Les besoins en formalismes adaptés à l'information spatiale

En 2003, à mon entrée dans l'équipe SI du Cemagref, des recherches étaient déjà initiées sur les formalismes de conception des SI spatiaux (notamment avec la thèse de doctorat préparée par André Miralles). Ce thème présente un grand intérêt pour la modélisation de SI environnementaux (Miralles, Pinet & Bedard, 2010; Papajorgji et al., 2010). J'ai donc étudié les formalismes de conception pour les SI spatiaux afin de voir quelles nouvelles pistes de recherche il était important de développer en lien avec les travaux de l'équipe. Ce chapitre présente une version réactualisée de l'étude des formalismes et décrit comment ses conclusions m'ont amené à positionner mes recherches sur la qualité des SI agri-environnementaux, et plus spécifiquement sur la représentation et le contrôle des contraintes d'intégrité spatiale. Nous soulignons dans ce chapitre que les formalismes existants ne permettaient pas (ou peu) la spécification de contraintes d'intégrité.

Les formalismes ER¹⁸ et OO¹⁹ sont depuis déjà longtemps les modes de représentation de prédilection des informaticiens, mais aussi, dans des versions plus « dépouillées », de certains thématiciens. En pratique, ils les utilisent le plus souvent pour réaliser des modèles de données et de classes. Depuis 20 ans, plusieurs équipes de recherche ont étudiées la personnalisation des formalismes ER puis OO pour faciliter la modélisation des données spatiales. L'information spatiale étant une donnée complexe, l'objectif des chercheurs est de simplifier (de clarifier) sa représentation dans les modèles de données. Cet objectif général est motivé de manière assez efficace dans la littérature (Bédard & Paquette, 1989; Kusters, Pagel & Six, 1996; Lbath, 1997; Shashi et al., 1997). Par exemple la Figure 2 et la Figure 3 tirées de (Lbath, 1997) montrent que la structure complexe des objets géographiques peut être simplifiée par l'utilisation de pictogrammes dans les modèles. Ces pictogrammes masquent les détails de la structure des objets spatiaux. Cet exemple illustre bien la philosophie générale des travaux dans le domaine.

Les objectifs des formalismes pour l'information spatiale qui ont été proposés sont de fournir des concepts et des notations spécifiques utilisables dans les modèles conceptuels afin de faciliter :

1. la représentation des types d'objets spatiaux de base (point, ligne, polygone, multi-points, multi-lignes, etc.) ; il s'agit d'un but commun à tous les formalismes proposés (illustré sur la Figure 3),
2. l'énumération de relations spatiales entre les objets (par exemple « une maison est spatialement incluse dans zéro ou un seul lotissement ») (Belussi, Negri & Pelagatti, 2006; Borges, Davis & Laender, 2001; Gubiani & Montanari, 2008; Kang et al., 2004; Lbath, 1997; Pantazis & Donnay, 1996; Parent et al., 1997; Pinet & Lbath, 2003; Salehi et al., 2007; Wang, Fierbinteanu & Maekawa, 2003) ; ce qui peut être vu dans le contexte des bases de données, comme la spécification de contraintes d'intégrité,
3. la description des évolutions des objets spatiaux dans le temps (les changements de forme, les objets mobiles, les trajectoires, etc.) (Brodeur, Bédard & Proulx, 2000; Claramunt & Parent, 2003; Claramunt et al., 1999; Kusters, Pagel & Six, 1996; Parent et al., 1997; Rocha,

¹⁸ ER : Entité-Relation

¹⁹ OO : Orienté-Objet

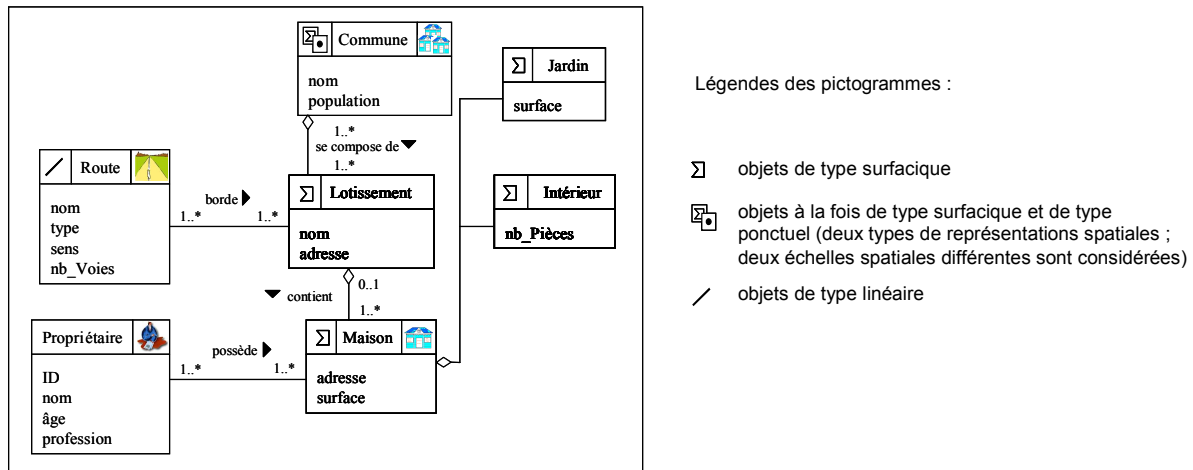


Figure 3. Un modèle représenté avec le formalisme proposé par (Lbath, 1997) (OMEGA) et décrivant les mêmes données spatiales que le modèle de la Figure 2 ; des pictogrammes indiquent la géométrie des objets (ponctuelle, linéaire, surfacique)

A noter que les objectifs 1-4 sont les plus traités. Toutes les propositions ont comme cadre d'utilisation directe la conception des modèles de données dans les BD transactionnelles. A noter néanmoins que certains travaux récents sont spécifiquement adaptés à la modélisation de l'information spatiale dans les entrepôts de données (Bédard et al., 2002; Malinowski & Zimanyi, 2008).

1.2 Historique

Depuis 20 ans, différentes extensions des formalismes ER et OO sont ainsi apparues :

- Fin 80-début 90 :
 - Modul-R (Bédard, Pageau & Caron, 1992; Bédard & Paquette, 1989) (*fondé sur le modèle ER*),
 - Congoo (Pantazis, 1994; Pantazis & Cornélis, 1996; Pantazis & Donnay, 1996) (*basé sur OOA²⁰ (Coad & Yourdon, 1991)*),
- Deuxième moitié des années 90 :
 - GeoOOA (Kosters, Pagel & Six, 1995, 1997) (*OOA*),
 - GISER (Shashi et al., 1997) (*ER*),
 - GeoER (Hadzilacos & Tryfona, 1997) (*ER*),
 - Mads (Parent, Spaccapietra & Zimanyi, 2006; Parent et al., 1997) (*OO*),
 - Pollen (Gayte et al., 1997) (*OMT²¹ (Rumbaugh et al., 1991)*),

²⁰ OOA : Object-Oriented Analysis

²¹ OMT : Object-Modeling Technique

- Omega (Laurini, 2001; Lbath, 1997) (*UML*),
 - GeoOM (Tryfona, Pfoer & Hadzilacos, 1997) (*OMT*),
 - STER (Tryfona et al., 1999) (*ER*),
 - Perceptory (Bédard, 1999; Bédard et al., 2004) (*UML*),
 - Extended Spatiotemporal UML (Price, Tryfona & Jensen, 1999; Price, Tryfona & Jensen, 2000) (*UML*),
 - UML-GeoFrame (Lisboa Filho & Cirano, 1999; Lisboa Filho, Cirano & Beard, 1998; Lisboa Filho & Lochpe, 2008) (*UML*),
 - OMT-G (Borges, Laender & Davis, 1999; Borges, Davis & Laender, 2001) (*OMT*),
- A partir de 2000 :
- Icons for GIS (Tveite, 2001) (*ER et UML*),
 - Semantics data model of spatiotemporal database (Sözer et al., 2008; Yazici, Zhu & Sun, 2001) (*UML*),
 - T-Omega (Pinet, 2002; Pinet & Lbath, 2003) (*UML*),
 - Conceptual framework for Spatiotemporal data modeling (Wang, Fierbinteanu & Maekawa, 2003) (*OO*),
 - ST USM (Khatri, Ram & Snodgrass, 2004, 2006) (*ER*),
 - GEOUML (Belussi, Negri & Pelagatti, 2004; Pegoraro & Velluto, 2007),
 - Profil SIG-Objecteering (Miralles, 2004, 2006; Miralles et al., 2009) (*UML*),
 - ChronoGeoGraph (Gubiani & Montanari, 2008) (*ER*)

Le Tableau 2 de la Section 1.5 de ce chapitre récapitule les objectifs de modélisation atteints pour chaque formalisme.

Peu après l'émergence du thème de recherche fin 80-début 90, de nombreux travaux furent initiés dans la deuxième moitié des années 90. Le nombre relativement important de propositions faites au cours de cette dernière période s'explique principalement par la nouveauté du sujet, mais aussi par l'effervescence d'alors autour des méthodes OO, ainsi que l'apparition d'UML (1995-97). Après 2000, on assiste à une « rationalisation » du domaine. Les formalismes pour l'information spatiale deviennent un axe de recherche important pour certaines équipes, alors que d'autres ne poursuivent pas leurs investigations sur le sujet. Les équipes actuellement très actives sur le thème, sont plus spécifiquement celles qui sont arrivées à offrir et à maintenir une solution logicielle solide mettant en œuvre leurs propositions de recherche. Il s'agit principalement :

- de l'équipe « bases de données spatiales » du CRG²² (Bédard, 2009) (*Perceptory*)

²² Centre de Recherche en Géomatique de l'Université Laval

- du laboratoire de bases de données de l'EPFL²³ en collaboration avec l'Université de Lausanne et l'ULB²⁴ (Parent, Spaccapietra & Zimanyi, 2009) (*Mads*),
- du Département d'informatique de l'UFV²⁵ (Lisboa Filho, 2009) (*UML-geoFrame*),
- du Cemagref (Miralles, 2009) (*Profil SIG-UML*).

Les outils développés permettent principalement de créer les modèles de données spatiales à l'aide d'une interface graphique adaptée, et de générer du code - par exemple des scripts de création de tables de base de données, etc. Les paragraphes suivants résument dans les grandes lignes, l'historique des activités de ces laboratoires qui restent aujourd'hui actifs dans le domaine et qui proposent des solutions logicielles fiables.

Y. Bédard²⁶ et F. Paquette initièrent le domaine de recherche à la fin des années 80 (Bédard & Paquette, 1989). Partant du constat qu'une modélisation détaillée de l'information spatiale rendait rapidement illisible les diagrammes ER, ils proposèrent les premiers d'utiliser des pictogrammes afin de décrire la géométrie des objets (cf. Figure 3). Ensuite, Y. Bédard expérimenta sur plusieurs projets son approche et travailla sur l'amélioration du formalisme proposé. Un premier outil de type Atelier de Génie Logiciel (AGL) nommé Orion fut développé. Il implémentait une nouvelle version du formalisme ER appelée MODUL-R (Bédard, Pageau & Caron, 1992). Dans la seconde moitié des années 90, Bédard et son équipe passent à l'OO en proposant la première version de l'outil Perceptory (Bédard, 1999). Cet outil était un module d'extension de MS Visio qui permettait de créer des modèles UML incluant les pictogrammes de MODUL-R. Depuis, les recherches et les développements autour de cet outil et du formalisme UML associé continuent. Perceptory connaît un franc succès qui illustre l'existence d'un besoin réel en formalisme dédié à l'information spatiale. (Bédard et al., 2004) indique plus de 300 téléchargement de ce gratuit par mois, ainsi qu'une utilisation dans plus de 40 pays. Perceptory a été utilisé dans de nombreuses applications (gestion des ressources naturelles, santé, archéologie, sport, etc.) (Bédard et al., 2004).

Les travaux sur le formalisme MADS (Modeling of Application Data with Spatio-temporal features) démarrèrent en 1995 dans le cadre d'un projet impliquant l'EPFL, l'Administration Cantonale Vaudoise, Texas Instrument Software et l'institut FORGE de l'Université de Lausanne (Parent, Spaccapietra & Zimanyi, 2009). En 1997, au terme du projet, les premières spécifications de MADS étaient terminées. Le laboratoire de bases de données de l'EPFL était le contributeur des propositions de recherche concernant le formalisme. MADS fut au fil du temps enrichi par le laboratoire pour traiter de manière efficace des problématiques spécifiques telles que la multi-représentation ou encore l'évolution temporelle des objets. Certaines de ces recherches furent menées en collaboration avec d'autres chercheurs (de l'Université Libre de Bruxelles, de l'Université de Nottingham,...). Un outil de type AGL a été développé dans le cadre d'un projet nommé COBALT au tout début des années 2000 (Parent, 2000). Il fut ensuite amélioré pour devenir MADSTRA. Ce développement fut réalisé dans le cadre du projet européen MurMur (2000-2002) qui concernait la multi-représentation (MurMur, 2009; Parent, Spaccapietra & Zimanyi, 2009). Le projet impliquait : l'EPFL, l'Université de Lausanne, le Cemagref (centre de Grenoble), l'ULB, l'IGN, Star. En 2006, sortit un ouvrage sur MADS écrit par les trois principaux chercheurs ayant contribué au formalisme au fil du temps : C. Parent (Université de Lausanne), S. Spaccapietra (EPFL) et E. Zimanyi (ULB) (Parent, Spaccapietra & Zimanyi, 2006). Il présentait de manière très détaillée le formalisme ainsi que des expérimentations

²³ Ecole Polytechnique Fédérale de Lausanne

²⁴ Université Libre de Bruxelles

²⁵ Université Fédérale de Viçosa

²⁶ Professeur au Centre de Recherche en Géomatique - Université Laval, Québec, Canada

réalisées dans le cadre de MurMur. MADS est depuis le début un formalisme reposant sur le paradigme ER étendu à certains concepts de l'OO (l'héritage par exemple). Comme indiqué dans (Parent, Spaccapietra & Zimanyi, 2009) et (Parent, Spaccapietra & Zimányi, 2006), MADS a servi pour la conception de modèles de données dans différents domaines : gestion pétrolière (en Colombie), gestion de réseau urbain d'eau (Genève), surveillance de rivières, gestion des ressources en eau (canton de Vaud), risques liés aux avalanches. L'application concernant le canton de Vaud fait notamment ressortir que MADS permet de grandement clarifier les modèles. Une réduction de 22% du nombre d'entités et de relations par rapport à un modèle ER traditionnel, a été estimée pour cette application (sans perte sémantique) (Parent et al., 1998).

A la fin des années 90, des recherches dirigées par Jugurta Lisboa à l'UFV firent émerger le formalisme GeoFrame basé sur UML (et qui sera rebaptisé par la suite UML-GeoFrame) (Lisboa Filho & Cirano, 1999; Lisboa Filho, Cirano & Beard, 1998). Après 2000, un outil logiciel fut développé ; aujourd'hui sa dernière version en date consiste en une implémentation de GeoFrame dans un générateur de code existant appelé ArgoUML (Victor de Freitas et al., 2005). ArgoUML est un outil open source assez connu, ancêtre de l'AGL commercial Poseidon. Il est simple d'utilisation tout en étant relativement performant pour la génération de code. La version de ArgoUML étendue au spatial a été appelée ArgoCASEGEO (Victor de Freitas, Lisboa Filho & Vilela, 2005). Elle est téléchargeable sur (Lisboa Filho, 2009). Les possibilités de modélisation offertes par GeoFrame restent encore limitées mais le formalisme est régulièrement amélioré. Les propositions sont très largement inspirées des concepts introduits par Perceptory et MADS pour représenter les types d'objets spatiaux et leur évolution temporelle. L'approche semble ne pas avoir encore été très expérimentée dans des domaines applicatifs.

C'est principalement aux alentours de 1995 que le Cemagref a commencé à s'intéresser au génie logiciel. Plus particulièrement dans le cadre d'un projet en hydraulique appelé X-NT (Cemagref, 2009). L'objectif était de développer une boîte à outils logicielle dans le domaine de la modélisation hydraulique à surface libre. La version ER de l'AGL Objecteering (Softteam, 2009) était employée dans le projet. UML fut par la suite utilisé dans différents autres projets du Cemagref (avec les nouvelles versions de l'AGL). Au fil du temps, le besoin d'avoir un formalisme plus adapté à l'information spatiale s'est fait sentir. La thèse d'André Miralles (Cemagref Montpellier) a démarré sur le sujet au début des années 2000. Ces travaux proposèrent une méthode de prototypage rapide pour les applications SIG. La méthode a été intégrée dans Objecteering et s'appuie sur un profil SIG-UML qui est assez proche de Perceptory. Cette approche basée sur l'ingénierie des modèles (Miralles et al., 2009) a été testée au Cemagref lors de la conception d'une application pour la maintenance des digues, et dans le cadre d'un projet de conception d'observatoires de pratiques agricoles. Les contributions apportées par cette thèse s'appliquent bien sûr, aussi, à d'autres domaines que ceux du Cemagref. La thèse a été soutenue en 2006 (Miralles, 2006). L'implémentation de la solution dans Objecteering, un AGL particulièrement puissant notamment en termes de génération de code, constitue un avantage majeur.


1.3 Synthèse des principales propositions existantes

Des comparatifs des différents formalismes sont régulièrement réalisés dans des articles scientifiques de synthèse ou des thèses ; voir par exemple (Claramunt, Coulondre & Libourel, 1997; Friis-Christensen, Tryfona & Jensen, 2001; Lbath, 1997; Miralles, 2006; Miralles, Pinet & Bedard, 2010; Pinet, 2002). Il ressort de ces comparatifs que les différentes méthodes proposent souvent des solutions de modélisation qui sont très proches. Plutôt que de refaire un comparatif des méthodes, nous présentons ici une synthèse de ce qui nous semble être les grandes avancées du domaine. Elles concernent trois sujets que l'on retrouve souvent dans la littérature du domaine : les représentations des objets spatiaux, leurs relations topologiques, et leurs évolutions dans le temps.

1.3.1 Représentation des objets spatiaux

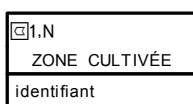
Les objets spatiaux stockés dans les SI sont de différents types, par exemple :

- Point, Ligne, ou Polygone,
- Multi-points, Multi-lignes, ou Multi-polygones,
- Ensemble d'objets de types hétérogènes, etc.

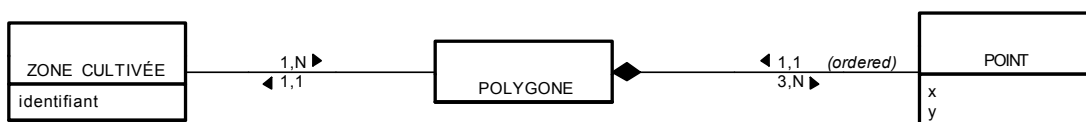
Il est assez clair que la simplification de la représentation de ces types de données dans les modèles, constitue un apport important des différents formalismes dédiés au spatial. Plus précisément, l'objectif des formalismes est de masquer le détail de la structure des objets spatiaux. Par exemple, sur le modèle de la Figure 3, le pictogramme  associé à la classe Commune indique qu'une commune est représentée par un polygone ou un point dans le SI sans pour autant donner de détails précis sur l'implémentation de ces objets spatiaux. Comme remarqué par (Miralles, 2006), ce type de modèles peut donc être vu comme un niveau d'abstraction au-dessus des modèles ER ou UML "classiques". Ce niveau d'abstraction moins détaillé facilite l'appropriation des diagrammes par les différents acteurs d'un projet. Lors du passage à une phase plus proche de l'implémentation, il est ensuite possible de transformer ces diagrammes qui utilisent des pictogrammes, en modèles ER ou UML classiques. Les stratégies de transformation sont multiples comme l'illustre l'exemple de la Figure 4. Sur la Figure, le pictogramme de la classe indique qu'une zone cultivée est composée de 1 à N polygones (conformément à la sémantique des notations Perceptory).

Dans la stratégie 1, la classe Polygone est reliée à une classe Point par une relation de composition ; il s'agit de points du contour du polygone qui seront par exemple utilisés pour stocker l'objet dans une base de données. Ainsi chaque polygone est associé à un ensemble ordonné de points de référence. La relation de composition indique qu'un point n'est rattaché qu'à un seul polygone. Si le polygone est détruit du système, tous ses points le sont eux aussi. Même si les contours de deux polygones différents ont un point (x,y) de référence en commun, il s'agira de deux instances différentes de la classe Point. Elles auront toutes les deux les mêmes coordonnées (x,y).

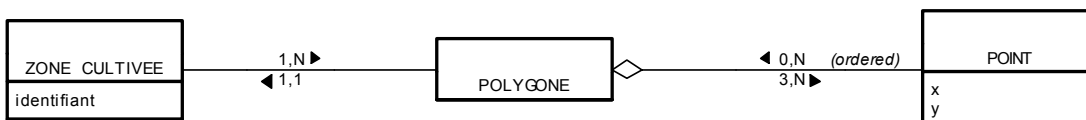
Dans la stratégie 2, c'est une relation d'agrégation qui relie la classe Polygone et la classe Point. Si deux polygones partagent un même point (x,y) de référence, alors il est possible de le représenter par une seule instance de la classe Point. Si un polygone est détruit du système, tous les points associés ne seront pas forcément détruits.

Classe d'un modèle (formalisme Perceptory) :Exemple de 2 stratégies différentes de représentation en UML :

Stratégie 1 :



Stratégie 2 :

**Figure 4. Deux exemples de transformation d'un modèle Perceptory en UML**

Dans les formalismes dédiés au géographique, des pictogrammes différents sont proposés pour chacun des types d'objet de base, par exemple : point, ligne, polygone. Certains formalismes proposent de combiner ces pictogrammes afin de modéliser des géométries plus complexes au sein des classes. Le Tableau 1 indique par exemple les différentes combinaisons possibles avec le formalisme Perceptory. Dans ce tableau, on retrouve dans la partie « Multiple geometries » le type de pictogrammes utilisé sur la Figure 4.




























| Grammar | Notation | Description and examples |
|---|--|--|
| Simple geometry | | |
| One pictogram with implicit 1,1 default multiplicity (not written). | Ex.  Ex.  Ex.  | Every instance is represented by one and only one simple geometry. Ex. A 'Fence' is represented by one and exactly one 1D geometry. |
| Aggregate geometry (AND) | | |
| Simple aggregate has a simple pictogram with 1,N multiplicity. | Ex.  1,N Ex.  1,N | Each instance is represented by an aggregate of shape of <i>same dimension</i> . Ex. An 'Orchard' is an aggregate of points (not a polygon). |
| Complex aggregate has different shapes in the same pictogram where shape order has no incidence. | Ex.  Ex.  Ex.  | Each instance is represented by an aggregate of shape of <i>different dimensions</i> . Ex. 'Large rivers' are composed of lines and polygons. <i>An aggregate of aggregates is an aggregate (i.e. no need of 1,N multiplicity).</i> |
| Alternate geometry (XOR) | | |
| Two or more pictograms on the same line. | Ex.  Ex.  Ex.  1,N  1,N Ex.  1,N | Each instance is represented by a geometry or another but not both (the exclusive OR). The proposed geometries can be simple and/or aggregates. Ex. 'Building' can be represented by a point if it is smaller than a fixed value or by a surface if it is larger. |
| Multiple geometries | | |
| Two or more pictograms on different lines | Ex.   Ex.  1,N  | Each instance is represented by each geometry but usually only one is used at a time. Ex. A 'Town' may have two geometries, a surface delimited by its boundary and a point representing downtown. <i>All geometry variations can be used within multiple geometries.</i> |
| Complicated geometry | | |
| Exclamation mark pictogram <i>Rarely used as it is facultative</i> |  | When the analyst feels that although it is possible, expressing the geometry with previous pictograms becomes too complicated. Description is in the repository. Ex. Bus Network made of complex aggregations of roads (multiple lines) + bus stops (points or polygons) + parking lots (single or multiple polygons). |
| Any possible geometry | | |
| Wildcard pictogram <i>Rarely needed</i> |  | All geometries are possible, without restriction. Ex. Historical Feature as a point (ex. statue), line (ex. historic street), polygon (ex. park), polygon + line (ex. historic place with adjacent streets), set of polygons (ex. group of buildings), etc. |
| Temporarily undefined geometry | | |
| Question mark pictogram. <i>Rarely used.</i> |  | The geometry is temporarily unknown. Will be replaced by meaningful pictograms. |
| Grammar | Notation | Description and examples |
| Facultative geometry | | |
| Add a minimum multiplicity of 0 and a maximum multiplicity after the pictogram. | Ex.  0,N Ex.  0,1 Ex.  0,1 | Only certain instances have a geometry. For example, all buildings are in the database, but only public ones have a position and shape and will appear on maps. <i>All the above geometry variations can be facultative.</i> |
| Derived geometry | | |
| Italicize the pictogram (to remind the slash used in UML). | Ex.  Ex.  Ex.  Ex.  | The geometry is obtained from the processing of other geometries or attributes. Ex. Country polygons can be derived from state polygons. <i>All the above geometry variations can be derived, each pictogram independently from the other.</i> |

Tableau 1. Exemple des combinaisons de pictogrammes possibles dans le formalisme Perceptory (tableau issu de (Bédard et al., 2004))

Ce principe de pictogrammes se retrouve dans les formalismes pour modéliser la multi-représentation spatiale (Balley, Parent & Spaccapietra, 2004; Davis & Laender, 1999; Proulx, Larrivée & Bédard, 2002; Spaccapietra, Parent & Zimányi, 2008; Vangenot, 2001), les objets spatiaux vagues (Miralles, 2006; Shu et al., 2003; Sözer et al., 2008; Yazici, Zhu & Sun, 2001), ou les structures de réseaux (Borges, Davis & Laender, 2001; Kusters, Pagel & Six, 1996). Ainsi, par exemple, certains formalismes proposent un pictogramme « objet vague » spécifiable dans les classes d'objets. Ceci indiquera que les objets des classes ont une géométrie incertaine. Pour le cas de la modélisation des structures de réseaux, on pourra associer aux classes, des pictogrammes « nœud du réseau », etc.

1.3.2 Modélisation des relations topologiques

Un autre apport important des formalismes proposés concerne la représentation des relations topologiques entre les objets spatiaux. Les auteurs des propositions considèrent très souvent que la modélisation des relations topologiques offre une information sémantique importante. La modélisation de ces relations peut être vue comme une spécification de contraintes d'intégrité, le concepteur décrivant les relations topologiques que l'on doit forcément trouver dans le SI modélisé.

Dès l'apparition des premiers formalismes dédiés aux SIG, on trouve l'idée de spécifier les relations topologiques au sein des modèles (Kusters, Pagel & Six, 1995; Pantazis, 1994). La méthode que l'on trouve le plus souvent consiste à représenter ces relations topologiques dans les diagrammes ER ou OO, par des relations entre entités ou classes (Borges, Laender & Davis, 1999; Borges, Davis & Laender, 2001; Hadzilacos & Tryfona, 1997; Kusters, Pagel & Six, 1995; Lbath, 1997; Pantazis, 1994; Parent et al., 1997; Tryfona, Pfoser & Hadzilacos, 1997). La sémantique des relations changent un peu d'une proposition à une autre. Par exemple, CONGOO (Pantazis, 1994) et OMEGA (Lbath, 1997) considèrent deux types de relations topologiques de base : le voisinage et la superposition. Le voisinage est considéré comme étant une intersection des frontières de la géométrie des objets ; par exemple deux polygones adjacents sont voisins. La superposition indique une intersection entre les objets. Le voisinage et la superposition entre deux objets géographiques peuvent être autorisés ou interdits. La Figure 5 donne un exemple de modèle CONGOO. Deux classes sont décrites : Routes nationales et Maisons. Les routes sont des Objets Géographiques Simples de type Ligne (OGS-L) et les maisons sont des Objets Géographiques Simples de type Polygone (OGS-Pol). L'association « -S » indique que la superposition entre les objets des deux classes est interdite (« +S » indiquerait que la superposition est autorisée).

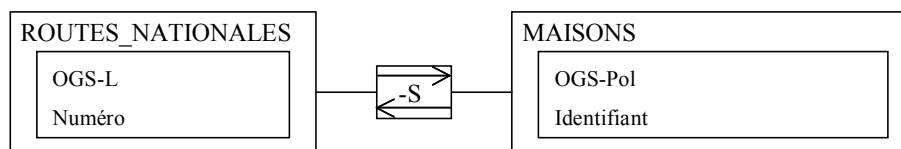


Figure 5. Exemple de diagramme CONGOO

Différents types de relations topologiques peuvent donc être spécifiés entre les classes sur un diagramme ER ou OO. D'autres approches recherchent des méthodes plus expressives pour représenter les relations topologiques. L'utilisation d'un outil logiciel dédié à la spécification des relations topologiques a été par exemple proposée en complément du formalisme Perceptory (Bédard et al., 2004) ; voir aussi (Cockcroft, 2004; Servigne et al., 2000). Pour notre part, nous avons utilisé le langage OCL (Object Constraint Language) pour modéliser précisément les relations topologiques entre les objets spatiaux décrits dans des modèles UML (Kang et al., 2004; Pinet & Lbath, 2003). Cette approche basée sur les contraintes d'intégrité, est décrite dans les chapitres suivants.

1.3.3 Modélisation de l'évolution temporelle des objets spatiaux

La description de l'évolution des objets spatiaux se fait principalement selon deux méthodes :



1. L'introduction de pictogrammes temporels dans les diagrammes. Ils permettent d'indiquer qu'un historique est associé à des attributs, des classes (ou entités), ou des relations. Les évolutions des valeurs des attributs, des instances de classes ou de relations devront donc être sauvegardées dans le SI. La Figure 6 présente un exemple de classe « Réseau Routier » modélisé en Perceptory. On trouve le pictogramme temporel  à droite du pictogramme décrivant la géométrie du réseau routier ( 1,N i.e. « de 1 à N lignes »). L'évolution spatiale des réseaux routiers peut se manifester par l'ajout ou la destruction de routes, ce qui change la géométrie globale des réseaux. Le pictogramme temporel indique que les différentes versions des réseaux routiers seront conservées (i.e. stockées) au sein de la classe. L'implémentation de cette classe devra donc permettre l'accès à n'importe laquelle de ces versions.



Figure 6. Exemple de pictogramme temporel du formalisme Perceptory (figure issue de (Bédard et al., 2004))

2. L'utilisation de relations temporelles entre les objets. Ces relations décrivent les liens d'évolution temporelle existant entre les différents objets spatiaux (fusion de plusieurs objets en un seul, subdivision d'un objet en plusieurs, etc.). Ces relations spécifiées sur les diagrammes d'entités ou de classes fournissent des informations importantes sur les changements d'états des objets. Les contributions des formalismes portent principalement sur la définition de relations d'évolution et de leur sémantique. La Figure 7 donne l'exemple de la relation de subdivision de MADS et GeoOOA. A une date donnée, une parcelle agricole peut être subdivisée en plusieurs autres parcelles.

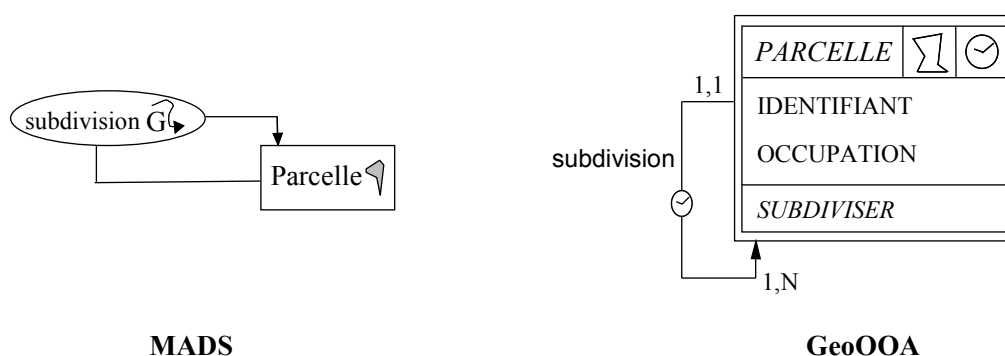


Figure 7. Exemple de relations temporelles avec MADS et GeoOOA

1.4 Positionnement des recherches initiées

A mon entrée au Cemagref, ma première préoccupation a été de m'intéresser aux formalismes de conception de SI spatiaux. A cette époque, les formalismes pour la conception des SI spatiaux avaient déjà atteint une certaine maturité. Les principales solutions existantes sont supportées par des outils. Le domaine de recherche reste très actif avec régulièrement de nouveaux articles sur le sujet. Au Cemagref, des recherches avancées sur les formalismes basés sur UML pour le spatial, ont été réalisées et utilisées notamment dans le cadre de la thèse de André Miralles (Miralles, 2006). Ces travaux ont aussi montré comment l'ingénierie dirigée par les modèles (i.e. l'approche Model-Driven Architecture (Kleppe, Warmer & Bast, 2003)) pouvait être appliquée avec ces formalismes. Une extension spatiale de l'AGL Objecteering a été proposée.

De plus en plus de sources différentes sont amenées à être intégrées dans les SI environnementaux afin d'y être analysées. Dans ce cadre, nous avons recherché comment enrichir les formalismes existants pour pouvoir modéliser facilement et précisément le niveau attendu de qualité des SI spatiaux modélisés. Ceci impliquait, entre autres, une description précise des relations topologiques attendues entre les objets spatiaux stockés dans le SI. La représentation des relations topologiques par des relations entre classes (cf. Section 1.3.2) offrait une méthode simple mais trop limitée pour représenter les règles de qualité complexes des SI agri-environnementaux ; par exemple une règle impliquant une condition comme « si un objet de la classe A contient spatialement un objet de la classe B, alors l'objet de A doit aussi être adjacent à un objet de la classe C ».

Nous avons étudié et proposé l'utilisation du langage OCL pour spécifier des règles complexes de qualité. Ces règles sont exprimées sous la forme de contraintes d'intégrité (Abiteboul, Hull & Vianu, 2000). OCL était parfaitement compatible avec les formalismes pour représenter l'information spatiale à condition qu'il soit lui-même étendu au spatial. Nous avons donc réalisé les travaux suivants :

1. Depuis 2003, nous avons étendu le langage OCL pour pouvoir exprimer des contraintes d'intégrité topologique impliquant des objets spatiaux. Nous avons proposé une solution pour générer le code qui permet de vérifier le bon respect de ces contraintes dans les bases de données.
2. Les formes de nombreux objets environnementaux sont incertaines car mal-connues, imprécises ou ne faisant pas l'objet d'un consensus. Un modèle innovant pour représenter de tels objets a donc été proposé ainsi qu'une extension spécifique d'OCL (depuis 2006).

3. Depuis 2008, nous travaillons sur l'expression des contraintes d'intégrité (notamment spatiale) dans les entrepôts de données, toujours avec OCL.

Après une synthèse concernant les méthodes opérationnelles pour modéliser les contraintes d'intégrité spatiale, les chapitres 3-4-5 décrivent dans l'ordre ces trois axes de recherche en précisant les expérimentations, réalisées ou en cours, dans le domaine de l'agriculture et de l'environnement.

1.5 Synthèse des fonctionnalités des différents formalismes

| | représentation des types d'objets spatiaux de base | spécification de relations topologiques | description des évolutions des objets spatiaux dans le temps | modélisation de la multi-représentation | description des objets aux limites vagues ou floues, aux positionnements incertains | modélisation de la structure des réseaux |
|-----------------------------|---|--|---|---|---|--|
| Modul-R | (Bédard, Pageau & Caron, 1992; Bédard & Paquette, 1989) | | | | | |
| Congoo | (Pantazis, 1994) | (Pantazis, 1994) | | | | |
| GeoOOA | (Kosters, Pagel & Six, 1995) | (Kosters, Pagel & Six, 1995) | (Kosters, Pagel & Six, 1996) | | | (Kosters, Pagel & Six, 1996) |
| GISER | (Shashi et al., 1997) | | | | | |
| GeoER | (Hadzilacos & Tryfona, 1997) | (Hadzilacos & Tryfona, 1997) | | | | |
| Mads | (Parent et al., 1997) | (Parent et al., 1997) | (Claramunt & Parent, 2003; Claramunt et al., 1999; Parent et al., 1997) | (Balley, Parent & Spaccapietra, 2004; Spaccapietra, Parent & Zimányi, 2008; Vangenot, 2001) | (Shu et al., 2003) | |
| Pollen | (Gayte et al., 1997) | | | | | |
| Omega | (Lbath, 1997) | (Lbath, 1997) | | | | |
| ST USM | (Khatri, Ram & Snodgrass, 2004) | | | | | |
| GeoOM | (Tryfona, Pfoser & Hadzilacos, 1997) | (Tryfona, Pfoser & Hadzilacos, 1997) | | | | |
| STER | (Tryfona et al., 1999) | | {Tryfona, 1999 #168} | | | |
| Perceptory | (Bédard, 1999) | | (Brodeur, Bédard & Proulx, 2000) | (Proulx, Larrivée & Bédard, 2002) | | |
| Extended Spatiotemporal UML | (Price, Tryfona & Jensen, 1999) | | (Price, Tryfona & Jensen, 1999) | | | |
| UML-Geoframe | (Lisboa Filho, Cirano & Beard, 1998) | | (Rocha, Edelweiss & Lochpe, 2001) | (Davis & Laender, 1999) | | |
| OMT-G | (Borges, Laender & Davis, 1999) | (Borges, Laender & Davis, 1999; Borges, Davis & Laender, 2001) | | | | (Borges, Davis & Laender, 2001) |
| Icons for GIS | (Tveite, 2001) | | | | | |

| | | | | | | |
|---|--------------------------------------|--|--------------------------------------|-----------------------------|---|------------------------------------|
| Semantics data model of spatiotemporal database | (Yazici, Zhu & Sun, 2001) | | | | (Sözer et al., 2008; Yazici, Zhu & Sun, 2001) | |
| T-Omega | (Pinet, 2002) | (Kang et al., 2004; Pinet & Lbath, 2003) | | | | |
| Conceptual framework for Spatiotemporal data modeling | (Wang, Fierbinteanu & Maekawa, 2003) | (Wang, Fierbinteanu & Maekawa, 2003) | (Wang, Fierbinteanu & Maekawa, 2003) | | | |
| GEOUML | (Belussi, Negri & Pelagatti, 2004) | (Belussi, Negri & Pelagatti, 2006) | | | | (Belussi, Negri & Pelagatti, 2004) |
| Profil SIG-Objecteering | (Miralles, 2004) | | (Miralles, 2006) | | (Miralles, 2006) | |
| ChronoGeoGraph | (Gubiani & Montanari, 2008) | (Gubiani & Montanari, 2008) | (Gubiani & Montanari, 2008) | (Gubiani & Montanari, 2008) | | |

Tableau 2. Buts atteints par les différents formalismes proposés ; les cellules indiquent des références qui se rapportent à ces objectifs

2 Synthèse des méthodes opérationnelles pour spécifier les contraintes d'intégrité spatiale

2.1 Introduction

Le contrôle des contraintes d'intégrité permet d'assurer la qualité des données et la cohérence d'un système d'information. Les données qui ne respectent pas ces contraintes sont considérées comme erronées. Il est donc important d'ajouter aux modèles de bases de données la spécification de ces contraintes. Les auteurs de (Abiteboul, Hull & Vianu, 2000) précisent que cette spécification offre un cadre pour ajouter de la sémantique aux modèles : "on incorpore des contraintes d'intégrité (en anglais *integrity constraints*), c'est-à-dire des propriétés supposées être satisfaites par toutes les instances d'un schéma de base de données."

Sur ce point, l'objectif de nos travaux était de fournir des méthodes pour modéliser (de manière expressive) les contraintes d'intégrité des bases de données spatiales, ainsi que mettre en place des outils pour les contrôler. Un challenge est de trouver un compromis entre le pouvoir d'expression offert par une méthode et sa complexité d'utilisation ; en d'autres termes, essayer d'exprimer le plus clairement possible les contraintes sans être trop limité dans les types de contraintes modélisables.

Certains formalismes décrits au Chapitre 1 supportent la spécification de relations topologiques entre des objets spatiaux, mais ils ne permettent pas de modéliser des contraintes d'intégrité spatiale complexes. Dans la littérature, il a été proposé des solutions spécifiques pour décrire les contraintes d'intégrité des bases de données spatiales (Cockcroft, 2004; Servigne et al., 2000; Ubada & Egenhofer, 1997). Ces propositions fournissent d'excellents résultats pour l'expression de contraintes d'un niveau de complexité simple ou moyen, et peuvent être utilisées facilement par des non-informaticiens. Dans nos travaux, nous avons recherché des méthodes plus expressives à destination des informaticiens, offrant la possibilité de spécifier à la fois des contraintes sur les données alphanumériques et sur les données spatiales. Utiliser un seul formalisme pour ces deux types de données permet ainsi de modéliser sous une même forme toutes les contraintes d'intégrité d'une base de données.

Des langages ou des méthodes offrent déjà la possibilité de décrire de manière déclarative et expressive des contraintes sur des données purement alphanumériques. C'est par exemple le cas d'OCL ou encore de la méthode B (Abrial, 1996; Mammar, Gervais & Laleau, 2006). Néanmoins, aucun de ces deux langages ne permettait de modéliser simplement les contraintes spatiales.

Avant de passer à l'état de l'art, nous présentons ci-dessous quelques définitions de base.

Les auteurs de (Bédard *et al.*, 2004) soulignent que les contraintes d'intégrité spatiale impliquent la sémantique et la géométrie des objets. D'une manière générale, les contraintes d'intégrité ont pour objectif de réduire les incohérences ou d'interdire des situations non désirées, et ainsi d'améliorer la qualité des données. Nous allons citer ici les principaux éléments définissant la qualité des données spatiales, et préciser lesquels de ces éléments nous avons cherché à améliorer grâce nos propositions. Par qualité des données, on entend l'adéquation des données avec certains critères (Devillers, 2004). Dans leur mémoire de thèse, (Ubada & Egenhofer, 1997) et (Puricelli, 2000) citent 7 éléments de qualité des données spatiales :

- La généalogie, "l'histoire des données", traite de la provenance des données et des moyens par lesquels elles ont été produites.
- La précision géométrique ou géographique, évalue le degré de conformité des données par rapport au monde réel en termes de position.

- La précision sémantique correspond à la précision des attributs non-spatiaux (une mesure par exemple) ; elle définit les différences entre les valeurs des attributs et leurs valeurs réelles.
- L'exhaustivité décrit si les objets du jeu de données représentent tous les objets du terrain.
- La cohérence logique se rapporte à la structure des données de la base. Par exemple, certains attributs appartiennent à un domaine de valeur particulier (latitude comprise entre 0 et 90°, etc.).
- La cohérence sémantique décrit la qualité avec laquelle les objets sont représentés par le modèle utilisé. Elle concerne essentiellement la pertinence et la signification des objets par la mesure de la "distance sémantique" entre les objets réels et leurs représentations.
- La précision temporelle traite des informations sur l'aspect temporel des données : leur date d'origine, leur fréquence de mise à jour, leur validité.

Les sources d'incohérence sémantique sont multiples. Par exemple, dans un SIG, une maison peut être située dans un lac pour différentes raisons : suite à une erreur de saisie, ou suite à une évolution temporelle non-enregistrée (si la rive du lac a été modifiée), ou encore suite à un problème de cohérence logique (le système ne permet pas d'indiquer que la maison est sur pilotis), ou enfin suite à un problème d'exhaustivité (une île a été oubliée). Les travaux de ce mémoire traitent principalement des contraintes d'intégrité qui ont pour but d'éviter les problèmes d'incohérence sémantique et logique.

2.2 Principaux moyens existants pour modéliser les contraintes d'intégrité

Cette section présente différentes solutions opérationnelles (supportées par des outils logiciels) pour exprimer les contraintes d'intégrité.

2.2.1 Outils pour la spécification de contraintes topologiques

Certains formalismes présentés au Chapitre 1 permettent de spécifier les relations topologiques entre objets. Des outils logiciels plus spécialisés ont aussi été proposés pour modéliser et contrôler l'intégrité topologique des SIG. Dans (Servigne et al., 2000; Ubeda & Servigne, 1997), les auteurs ont développé et utilisent une interface visuelle pour spécifier des contraintes topologiques entre les entités d'un SIG.

Avec l'interface proposée, le cheminement pour spécifier une contrainte est le suivant :

1. Sélectionner le premier groupe d'objets géographiques.
2. Sélectionner le second groupe d'objets géographiques.
3. Sélectionner la relation topologique.
4. Fixer la contrainte. Celle-ci précise que la relation entre les types d'objets sélectionnés est "interdite", ou encore qu'elle peut apparaître "exactement N fois" ou "au moins N fois" ou "au plus N fois".

Par exemple, nous pouvons définir avec l'outil la contrainte suivante entre *Route* et *Bâtiment* : les routes ne traversent pas les bâtiments (*Route-Traverse-Bâtiment-Interdit*). Il est aussi possible d'indiquer qu'une section cadastrale est dans exactement une commune (*SectionCadastrale-Dans-*

Commune-Exactement 1 fois). La Figure 8 présente l'interface de l'outil. La contrainte spécifiée sur cette figure est qu'une route ne doit pas être à l'intérieur ou traverser un bâtiment.

Dans (Cockcroft, 1998, 2001; Cockcroft, 2004), l'auteur présente un outil proche de celui de (Servigne et al., 2000; Ubeda & Servigne, 1997). La différence principale réside dans le fait que les relations topologiques spécifiées peuvent dépendre de conditions sur la valeur des attributs. Par exemple, il est possible de modéliser la contrainte suivante : une canalisation ne doit intersecter une valve que si le diamètre de cette canalisation est supérieur à 14 centimètres. La Figure 9 montre en effet que l'interface de l'outil permet d'associer des conditions de valeurs aux attributs. Csory est un autre outil logiciel très intéressant développé cette fois par l'Université Laval permettant de spécifier les relations topologiques d'une façon assez semblable (Bédard et al., 2004).

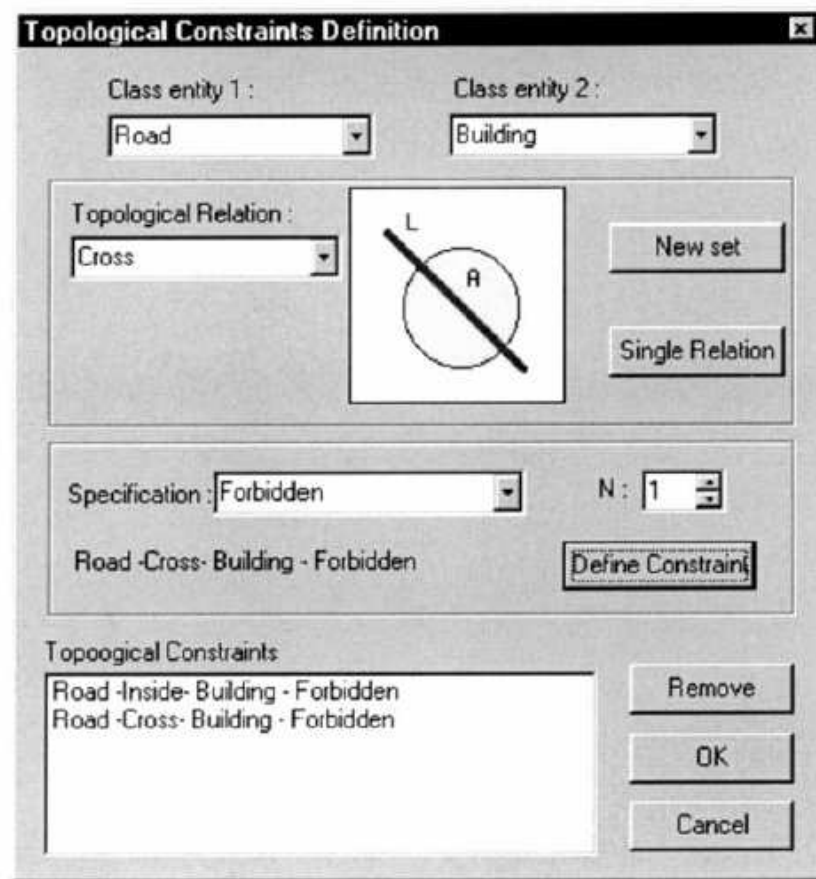


Figure 8. Interface de l'outil proposé par (Servigne et al., 2000; Ubeda & Servigne, 1997)

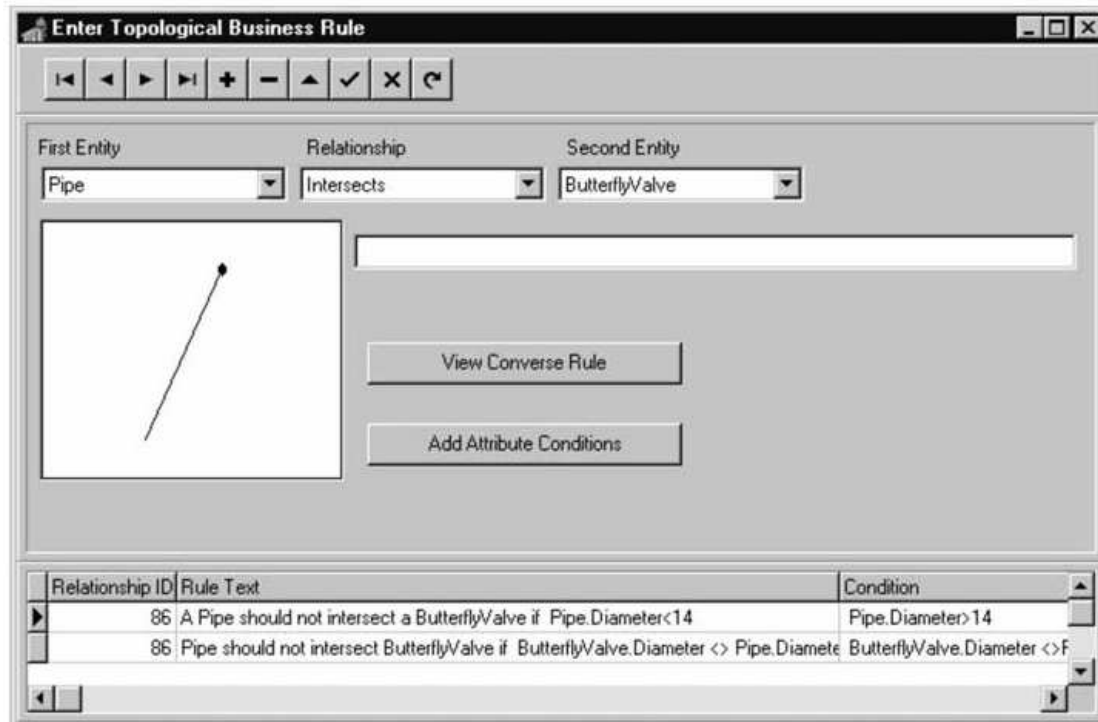


Figure 9. Interface de l'outil de (Cockcroft, 1998, 2001; Cockcroft, 2004)

Ces différentes méthodes permettent de spécifier de façon simple des contraintes dans les SIG. Elles permettent d'exprimer des contraintes topologiques plus complexes que celles modélisables avec les formalismes pour les SIG (présentés au Chapitre 1). Néanmoins, certaines contraintes complexes ne sont pas gérées. C'est par exemple le cas des deux contraintes suivantes :

- Pour qu'un lotissement d'habitation soit classé comme « zone bâtie à risque élevé d'inondation », il faut qu'au moins la moitié de ses habitations résidentielles se trouvent dans une « zone inondable ».
- Si un objet de la classe A contient spatialement un objet de la classe B, alors l'objet de A doit aussi être adjacent à un objet de la classe C.

2.2.2 Formalismes et langages expressifs pour la spécification de contraintes

Nous présentons ici les formalismes et les langages particulièrement expressifs pour spécifier des contraintes d'intégrité.

Pour définir des contraintes d'intégrité expressives, l'utilisation de "triggers" (déclencheurs) peut être envisagée. La mise en place de ces triggers est un objet de recherche important dans le domaine des bases de données actives (Widom & Ceri, 1995). Ces triggers écrits par exemple en SQL (voire en SQL spatial) se déclenchent automatiquement lors d'une mise-à-jour de la base de données pour contrôler une contrainte. Cependant, l'écriture directe des triggers reste lourde. SQL n'est pas un langage spécifiquement dédié aux contraintes. De plus, les triggers portent sur le schéma physique des bases et non sur le schéma conceptuel, ce qui peut complexifier leur écriture.

La méthode B est une méthode formelle de développement du logiciel, ayant connue quelques implémentations. Elle permet de spécifier des contraintes sur le comportement des logiciels. Il s'agit d'un formalisme déclaratif. Il permet de "parler du quoi, pas du comment" (Abrial, 1996). La méthode est applicable aux SI (Mammar, Gervais & Laleau, 2006). L'écriture de contraintes en B est dédiée à des utilisateurs confirmés. Aussi, B reste relativement déconnecté des langages usuels (notamment UML) et des Ateliers de Génie de Logiciel utilisés aujourd'hui dans la conception de bases de données. Aucune extension de B n'a été proposée pour le spatial.

OCL (Object Constraint Language) est un langage formel pour l'expression de contraintes appliquées à des diagrammes UML. Il fait maintenant partie intégrante du standard UML. C'est un langage déclaratif : il spécifie les conditions qui doivent être satisfaites. C'est un langage sans effet de bord : l'évaluation d'une contrainte ne modifie pas l'état du système. OCL a été conçu pour être à la portée des ingénieurs concepteurs de systèmes d'information. Le langage a été développé par un groupe de scientifiques d'IBM autour de 1995 au cours d'un projet de modélisation d'entreprise. Le langage a été influencé par Syntropy, un langage orienté-objet (Cook & Daniels, 1994). Un nombre grandissant de concepteurs de systèmes d'information utilise ce langage de contraintes en complément des diagrammes de classes d'UML pour exprimer les invariants des modèles par exemple. Dans (Mandel & Cengarle, 1999) et (Akehurst & Bordbar, 2001), les possibilités d'OCL sont étudiées pour exprimer des requêtes ; dans ces articles, l'expressivité du langage en temps que langage de requêtes est alors étudiée. Les auteurs montrent qu'en considérant la notion de tuples dans OCL, le langage peut supporter l'algèbre relationnelle. Un nombre grandissant d'outils implémentent OCL. On trouve aujourd'hui des plug-ins spécifiques pour Eclipse, Rational Rose, ArgoUML, etc. (Klasse Objecten, 2009). Dans le contexte des bases de données, les contraintes OCL sont modélisées au niveau conceptuel (sur le schéma conceptuel), ce qui simplifie grandement leur expression. Le langage OCL peut être facilement intégré à n'importe lequel des formalismes OO ou ER utilisés pour la conception de SI spatiaux (présentés au Chapitre 1).

Nous introduisons ici les concepts généraux du langage OCL à l'aide d'exemples. Une contrainte OCL peut se rapporter à une classe (appelée le contexte de la contrainte). Elle représente une condition qui

doit être satisfaite pour toute instance d'une classe. Les exemples suivants sont basés sur le diagramme de classes UML de la Figure 10.

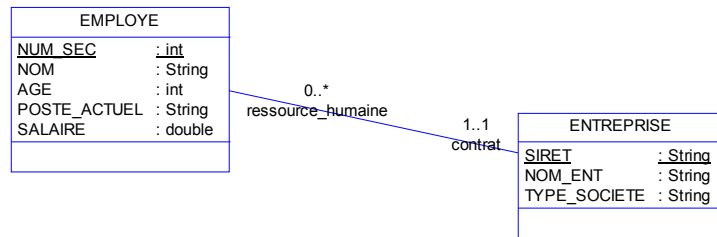


Figure 10. Les entreprises et leurs employés

La contrainte OCL suivante exprime le fait que tous les employés dont le poste actuel est chercheur perçoivent un salaire supérieur ou égal à 1500€.

```

context Employe inv:
  self.poste_actuel = 'chercheur' implies self.salaire >= 1500
  
```

Cette contrainte OCL est équivalente à la formule logique suivante :

```

∀self ∈ Employe,
  (self.poste_actuel = 'chercheur' → self.salaire >= 1500)
  
```

En OCL, *self* est toujours une variable-itérateur qui correspond à une instance du contexte (ici le contexte est la classe *Employe*). Une contrainte doit obligatoirement être respectée pour toute instance du contexte. *self.poste_actuel* et *self.salaire* sont deux attributs (cf. Figure 10). L'opérateur *implies* correspond à l'implication logique.

Navigation entre les classes. La navigation permet de fixer des contraintes impliquant des objets de classes différentes. Par exemple, la contrainte suivante signifie que les petites et moyennes entreprises (PME) n'ont pas plus de 499 salariés.

```

context Entreprise inv:
  self.type_societe = 'PME'
  implies
  self.ressource_humaine->size() <= 499
  
```

Cette contrainte OCL est équivalente à la formule logique suivante :

```

∀self ∈ Entreprise, ( self.type_societe = 'PME' →
  size(self.ressource_humaine) <= 499 )
  
```

L'attribut *type_societe* est défini dans la classe *Entreprise*, et le nombre de salariés s'obtient à partir des instances de la classe *Employe*. Une navigation est donc réalisée entre ces deux classes via l'association *ressource_humaine*. Ici, *self* correspond à une entreprise, et *self.ressource_humaine* retourne une collection contenant tous les employés de l'entreprise *self*. L'opération *size()* retourne la taille d'une collection (i.e. le nombre d'objets la composant).

Opérations sur les éléments d'une collection. La syntaxe d'une opération portant sur les éléments d'une collection est la suivante :

$\langle \text{collection} \rangle \rightarrow \langle \text{opération} \rangle ([\langle \text{itérateur} \rangle] \langle \text{expression} \rangle)$

où l'expression booléenne $\langle \text{expression} \rangle$ est évaluée pour chacun des éléments de la $\langle \text{collection} \rangle$. $\langle \text{itérateur} \rangle$ est une variable facultative qui sert à faire référence à l'élément courant dans l' $\langle \text{expression} \rangle$. $\langle \text{opération} \rangle$ peut être l'une des opérations suivantes (entre autres) :

1) *forAll* représente le quantificateur universel qui indique que l'expression booléenne doit être vraie pour toutes les instances de $\langle \text{collection} \rangle$. Par exemple, on peut spécifier le fait que tous les employés des établissements de recherche ont au moins 18 ans.

```
context Entreprise inv:
  self.type_societe = 'établissement de recherche' implies
  self.ressource_humaine->forAll(age>=18)
```

Si nous représentons la contrainte sous une forme logique, nous obtenons :

$$\forall \text{self} \in \text{Entreprise}, (\text{self.type_societe} = \text{'établissement de recherche'} \rightarrow (\forall x \in \text{self.ressource_humaine} (x.\text{age} \geq 18)))$$

2) *exists* représente le quantificateur existentiel qui indique que l'expression booléenne doit être vraie pour au moins un élément de $\langle \text{collection} \rangle$. Par exemple, on peut exprimer le fait qu'il existe (un ou) plusieurs chercheurs dans un établissement de recherche.

```
context Entreprise inv:
  self.type_societe = 'établissement de recherche' implies
  self.ressource_humaine->exists(poste_actuel='chercheur')
```

Sous une forme logique :

$$\forall \text{self} \in \text{Entreprise}, (\text{self.type_societe} = \text{'établissement de recherche'} \rightarrow (\exists x \in \text{self.ressource_humaine} \mid x.\text{poste_actuel} = \text{'chercheur'}))$$

3) *size* retourne le nombre d'éléments (i.e. le nombre d'objets) d'une collection.

4) *select* retourne un sous-ensemble de $\langle \text{collection} \rangle$ qui ne contient que les éléments qui vérifient l' $\langle \text{expression} \rangle$. Par exemple, on peut l'utiliser dans une contrainte pour sélectionner les employés majeurs en indiquant :

```
self.ressource_humaine->select(age>=18)
```

On utilise *allInstances* pour retourner toutes les instances d'une classe. *Employe.allInstances()* retourne par exemple la collection de toutes les instances de *Employe*. Il existe de très nombreuses autres opérations applicables sur les collections en OCL (OMG, 2007).

2.3 Conclusion

En résumé, nous avons d'un côté, différentes méthodes dédiées permettant de spécifier facilement des contraintes d'un niveau de complexité moyen dans les SIG (cf. Section 2.2.1), et de l'autre côté, des langages formels déclaratifs et expressifs pour décrire des contraintes complexes mais non spatiales (cf. Section 2.2.2).

Notre but était d'exprimer des contraintes d'intégrité spatiale complexes. Une solution pour atteindre notre objectif était donc d'intégrer la dimension spatiale dans un langage expressif de contraintes.

Notre analyse nous a amené à choisir OCL. La Section 2.2.2 a énuméré les nombreux avantages de ce langage. C'est un langage déclaratif conçu pour être à la portée des ingénieurs concepteurs de systèmes d'information. Un nombre grandissant d'informaticiens utilise ce langage de contraintes avec les diagrammes de classes d'UML et on trouve des plug-ins spécifiques pour Eclipse, Rational Rose, ArgoUML, etc. Dans le contexte des bases de données, les contraintes OCL sont modélisées au niveau conceptuel (sur le schéma conceptuel), ce qui facilite grandement leur spécification. OCL s'intègre très facilement aux formalismes utilisés pour la conception de systèmes d'information spatiaux (tels que ceux décrits au Chapitre 1).

Comme nous allons le présenter dans le chapitre suivant, nous avons donc proposé des extensions spatiales spécifiques pour ce langage. Plus précisément, nous offrons des solutions pour modéliser des contraintes spatiales intégrant des relations topologiques dans leur formalisation.

3 Modélisation de contraintes d'intégrité topologique impliquant des régions certaines

3.1 Introduction

L'état de l'art présenté dans le chapitre précédent nous a amené à choisir OCL pour modéliser les contraintes d'intégrité spatiale. Les nombreux avantages d'OCL énumérés en Section 2.2.2 nous ont amené à choisir ce langage. Ce mode de représentations s'intègre facilement aux formalismes utilisés pour la conception de systèmes d'information spatiaux (tels que ceux décrits au Chapitre 1). Nous avons proposé des extensions spatiales spécifiques pour OCL. Plus précisément, nous offrons des solutions pour modéliser des contraintes spatiales intégrant des relations topologiques dans leur formalisation.

Nous nous focalisons sur les contraintes d'intégrité des bases de données stockant des régions simples et composites. Nous entendons par région composite, une région faite de l'union de régions simples. Pour les contraintes entre deux régions simples, nous avons intégré des relations topologiques adaptées dans OCL. Pour les contraintes entre régions composites, nous proposons, d'une part, d'utiliser des opérations existantes d'OCL et d'autre part, nous introduisons de nouvelles fonctionnalités dans OCL pour spécifier plus directement ce type de contraintes. Nos propositions de langages sont intégrées dans un outil de génération de code SQL appelé OCL2SQL. Ce générateur produit automatiquement, à partir de contraintes spécifiées au niveau conceptuel, des mécanismes permettant de tester et de garantir l'intégrité d'une base de données.

Les travaux de ce chapitre ont été réalisés en grande partie dans le cadre de la thèse de Magali Duboisset préparée et soutenue au Cemagref (Duboisset, 2007). Historiquement, nous avons introduit des premières idées d'extension d'OCL pour exprimer des contraintes d'intégrité spatiale dans (Pinet, 2002). Nous avons brièvement présentées les idées de base de la mise en place d'un OCL pour les contraintes d'intégrité spatiale dans (Duboisset et al., 2005a, b; Duboisset, Pinet & Schneider, 2007; Pinet, Duboisset & Soullignac, 2007) puis de manière plus détaillée dans (Pinet et al., 2009a; Pinet, Duboisset & Schneider, 2009).

Le chapitre est organisé de la façon suivante. La Section 3.2 présente nos extensions spatiales d'OCL. La Section 3.3 concerne l'implémentation développée sous la forme d'un outil de type générateur de code. La Section 3.4 présente une étude de cas, détaillant l'expérimentation de la solution et l'illustrant sur plusieurs exemples de contraintes. Nous décrivons les tests réalisés dans le cadre du projet SIGEMO (Système d'Information sur la Gestion des Epanchages de Matières Organiques). La Section 3.5 conclut et précise les perspectives des travaux.

3.2 Extensions spatiales d'OCL

Afin de permettre une définition de contraintes d'intégrité spatiale à la fois expressives et déclaratives, notre proposition a consisté à effectuer des extensions d'OCL. Nous avons donc intégré des opérations topologiques à OCL.

Les SGBD actuels supportent principalement les relations topologiques entre régions issues des deux modèles « 9-Intersection Method » (9-IM) (Egenhofer & Herring, 1992) et « Calculus Based Method » (CBM) (Clementini & Di Felice, 1995). Afin d'être compatible avec les SGBD existants, nous avons proposé deux extensions d'OCL intégrant ces relations : OCL_{9IM} (Duboisset et al., 2005b; Pinet et al., 2009a; Pinet, Duboisset & Schneider, 2009; Pinet, Duboisset & Soullignac, 2007) et OCL_{CBM} (Duboisset et al., 2005a; Duboisset, Pinet & Schneider, 2007). Dans ce mémoire nous introduisons et

illustrerons l'esprit général de nos propositions en détaillant plus spécifiquement OCL_{9IM} . OCL_{CBM} suit exactement la même approche que OCL_{9IM} mais en se basant sur les relations topologiques de CBM.

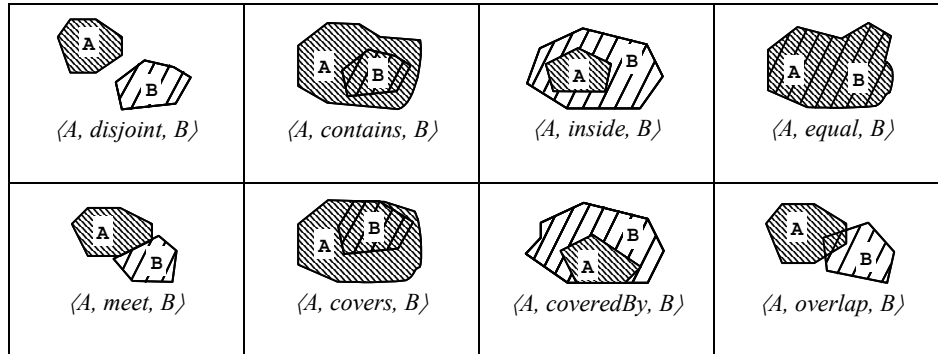


Figure 11. huit relations topologiques possibles entre deux régions simples (Egenhofer & Herring, 1992) (Egenhofer & Franzosa, 1991)

OCL_{9IM} est le langage OCL auquel nous avons ajouté les 8 relations du modèle 9IM entre régions simples (voir Figure 11).

Définition 2.1. Région (simple).

Une région (simple) est un ensemble de points connectés, sans trou, dans R^2 .

□

Le type *Region* est inclus dans le langage OCL et chaque relation topologique de 9IM est intégrée dans OCL sous la forme d'une nouvelle opération.

Définition 2.2. Opérations spatiales d' OCL_{9IM} .

La syntaxe générale de ces opérations spatiales est décrite ci-dessous.

$(geoA).opérationTopo(geoB) : Boolean$

où $opérationTopo \in \{ inside, contains, covers, coveredBy, disjoint, equal, overlap, meet \}$ et $geoA, geoB$ sont de type *Region*.

$geoA, geoB$ sont les paramètres de l'opération. Chaque opération retourne vrai ou faux (un booléen) suivant que la relation topologique entre $geoA$ et $geoB$ est vraie ou fausse.

□

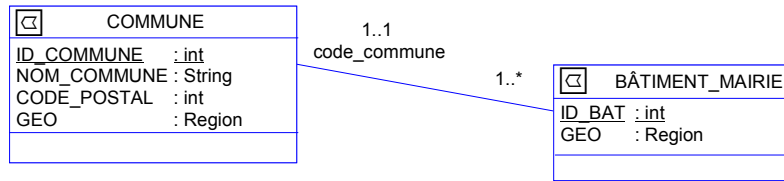


Figure 12. Les bâtiments de mairie et leur commune

Par exemple, les bâtiments d'une mairie sont généralement localisés dans la commune qui correspond à leur adresse. En se basant sur le modèle de la Figure 12, cette contrainte s'écrit comme suit en $\text{OCL}_{9\text{IM}}$:

```

context Bâtiment_Mairie inv:
  (self.geo) .inside      (self.code_commune.geo) or
  (self.geo) .coveredBy (self.code_commune.geo)
  
```

La contrainte doit être vérifiée pour chaque instance *self* de *Bâtiment_Mairie*. L'expression *self.code_commune* retourne l'instance de *Commune* associée à *self* par l'association *code_commune*.

Afin d'augmenter les possibilités de notre langage, nous nous sommes aussi intéressés aux relations topologiques entre régions composites. Nous entendons par région composite, une union de régions telle que définie ci-dessous.

Définition 2.3. Union de régions (région composite) (Claramunt, 2000).

Une union de régions UR est un ensemble fini, non vide, de régions simples R_1, \dots, R_n tel que :

- $R_i^\circ \cap R_j^\circ = \emptyset, \forall i \neq j$
- $\delta R_i \cap \delta R_j = \emptyset, \forall i \neq j$

où R° et δR sont respectivement l'intérieur et la frontière de R .

□

Nous avons vu que $\text{OCL}_{9\text{IM}}$ permet la spécification de contraintes entre des objets géométriques simples. Une région composite est un ensemble de régions simples. Par conséquent, les opérations standards d'OCL sur les collections (cf. Section 2.2.2) peuvent être utilisées en $\text{OCL}_{9\text{IM}}$ pour accéder aux différentes régions simples d'une région composite. Ainsi, les opérations standards comme *size*, *forall*, *exists* ou *select* peuvent être appliquées sur des attributs de type *Set(Region)* c'est-à-dire de type « région composite ».

Définition 2.4. Opérations d' $\text{OCL}_{9\text{IM}}$ permettant de modéliser des contraintes sur des régions composites.

La syntaxe générale des opérations sur les régions composites est :


```
geo -> opérationSurCollection(...)
```

où *geo* est de type *Set(Region)* et *opérationSurCollection* est une opération OCL sur les collections.

□

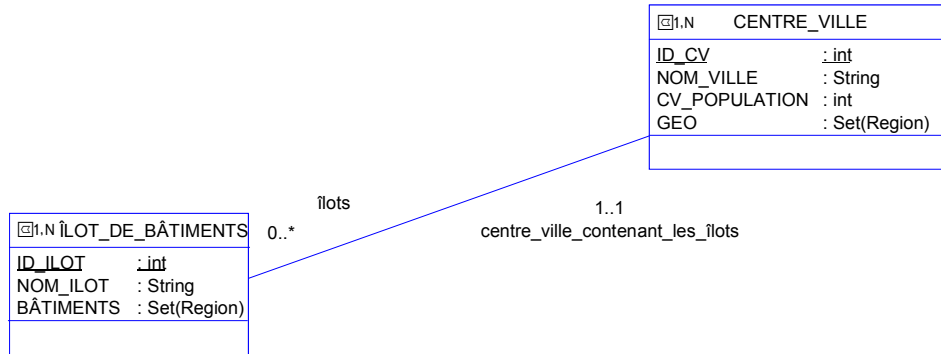


Figure 13. Les centres ville et leurs îlots de bâtiments ; on ne modélise que les îlots qui sont dans un centre ville

Ainsi, il devient possible d'exprimer des relations topologiques entre les régions composites. Par exemple, dans le modèle de la Figure 13, les attributs *bâtiments* de la classe *Îlot_de_Bâtiments* et *geo* de la classe *Centre_Ville* sont de type *Set(Region)*. Cela signifie, que chacun de ces attributs stocke une région composite. Chaque attribut *bâtiments* de la classe *Îlot_de_Bâtiments* stocke un îlot de bâtiments, c'est-à-dire un ensemble de bâtiments proches. Chaque valeur de cet attribut est l'ensemble des régions simples (i.e. une région composite) qui correspondent à l'îlot. Un centre ville est défini comme un ensemble de zones dans la ville. Ces zones sont stockées dans l'attribut *geo* de la classe *Centre_Ville*. L'association présente sur le diagramme de la Figure 13 indique que les zones d'un centre ville peuvent contenir des îlots de bâtiments. Ces classes sont illustrées sur un exemple à la Figure 14. Sur le dessin de cette figure, le centre ville est composé de 2 zones distinctes. Deux exemples d'îlots sont aussi indiqués, le premier îlot fait 6 bâtiments et le second en fait 4. Les opérations d'OCL sur les ensembles peuvent être appliquées aux attributs *bâtiments* et *geo* des classes, comme illustré ci-dessous.

Soit *I* un îlot et *C* un centre ville. Si *I* et *C* sont associés par *centre_ville_contenant_les_îlots* alors pour chaque bâtiment *b* de *I*, il doit exister une partie *c* du centre ville *C* tel que *b* est dans *c*. Cette contrainte s'exprime en OCL_{9IM} :

```

context Îlot_de_Bâtiments inv:
  self.bâtiments -> forAll( b |
    self.centre_ville_contenant_les_îlots.geo
      -> exists( c | (b).inside( c) ) )
  
```

Les attributs *geo* et *bâtiments* sont de type *Set(Region)*, et donc *b* et *c* sont de type *Region*.

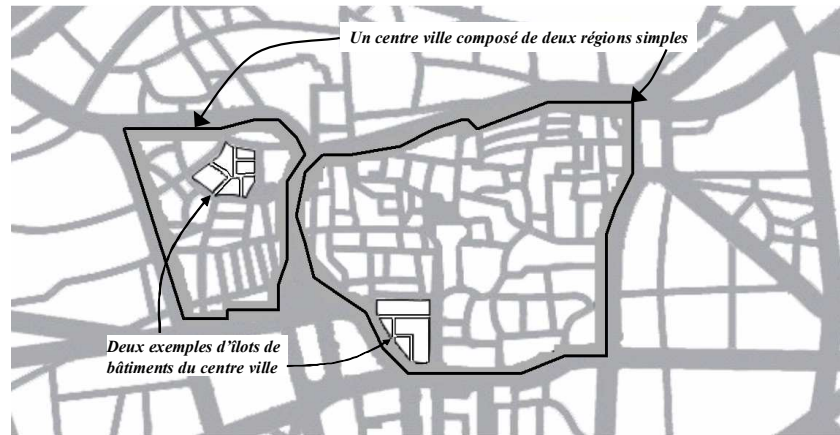


Figure 14. Exemple d'un centre ville et de ses îlots de bâtiments

Il est apparu assez rapidement que les contraintes sur les régions composites pouvaient devenir complexes à écrire. Nous avons donc cherché à faciliter l'expression de contraintes sur les régions composites. Nous avons intégré dans OCL, les deux modèles existants de relations topologiques entre régions composites :

- le modèle basé sur des adverbes proposés par (Claramunt, 2000). Il repose sur 9IM. Nous avons donc enrichi OCL_{9IM} avec ce modèle et obtenu le nouveau langage OCL_{ADV} .
- le modèle Topological Relationships for Composite Regions (TRCR) (Clementini, Felice & Califano, 1995) qui est basé sur CBM. Nous avons enrichi OCL_{CBM} avec ce modèle et obtenu le nouveau langage OCL_{TRCR} .

Nous avons réalisé des études visant à comparer les expressivités d' OCL_{9IM} , OCL_{ADV} , OCL_{CBM} et OCL_{TRCR} . Il ressort de ces études que OCL_{9IM} et OCL_{ADV} ont un pouvoir d'expression équivalent (Duboisset, 2007; Duboisset et al., 2005b; Pinet, Duboisset & Schneider, 2009). Ainsi, la différence entre les langages OCL_{9IM} et OCL_{ADV} se situe au niveau de la facilité d'expression des contraintes. OCL_{ADV} permet d'écrire certaines contraintes plus directement qu'avec OCL_{9IM} ; l'écriture avec OCL_{ADV} sera plus succincte. Il en va de même pour les propositions d'extensions spatiales d'OCL basées sur CBM : OCL_{CBM} et OCL_{TRCR} ont le même pouvoir d'expression (Duboisset, 2007; Duboisset et al., 2005a). Par contre, OCL_{9IM} est plus expressif que OCL_{CBM} (Duboisset, 2007).

3.3 Implémentation

Le but de cette partie est de présenter l'outil grâce auquel nous avons implémenté et expérimenté les langages proposés. Un enjeu majeur est de faciliter le passage entre la description conceptuelle des contraintes et leur implémentation dans les bases de données spatiales. Grâce à notre proposition, les contraintes OCL_{9IM} , OCL_{ADV} , OCL_{CBM} et OCL_{TRCR} sont traduites sous forme de triggers activés sur une base de données, ou bien sous la forme de requêtes permettant de tester l'intégrité de la base.

Un certain nombre d'implémentations d'OCL sont disponibles (Klasse Objecten, 2009) : par exemple, *Octopus*, qui génère des expressions Java à partir de contraintes en OCL, *Bold for Delphi* et *OCL2SQL* qui peuvent générer du code SQL. Afin de produire les mécanismes de vérification d'intégrité, nous avons choisi d'étendre l'outil *OCL2SQL* qui présente l'avantage d'être disponible en open source.

3.3.1 Présentation d'OCL2SQL

Le programme OCL2SQL est un générateur puissant écrit en Java (Demuth, 2005; Demuth & Hussmann, 1999; Demuth, Hussmann & Loecher, 2001; Demuth, Loecher & Zschaler, 2004; Drummond & Shearer, 2006; Pinet et al., 2009a). Il permet de générer automatiquement à partir d'une contrainte OCL :

- une requête SQL qui sélectionne toutes les données qui ne satisfont pas la contrainte ; l'exécution de cette requête permet donc de vérifier la cohérence de la base de données,
- un trigger ayant pour but de vérifier la base de données à chaque mise à jour ; pour ce faire, le trigger déclenche l'exécution de la requête SQL après chaque insertion, suppression et modification des données.

En fait, le trigger vérifie si la requête SQL produite par OCL2SQL retourne des tuples. Si ce n'est pas le cas, la mise à jour est définitivement acceptée, sinon elle est rejetée. Par cette technique, il devient impossible d'insérer des données qui violent la contrainte. L'explication de la structure générale des triggers générés est présentée dans (Demuth, Hussmann & Loecher, 2001).

Le principe de la génération de code est basé sur des patrons de conversion d'OCL vers SQL. OCL2SQL génère une requête correspondant à la négation de l'expression à respecter dans la contrainte. Cette requête retournera donc tous les enregistrements qui ne satisfont pas la contrainte. Par ailleurs, OCL2SQL est couplé à une interface graphique permettant la saisie, la sauvegarde et le contrôle syntaxique des contraintes OCL.

L'outil OCL2SQL permet de charger un fichier XMI (XML Metadata Interchange - format d'échange des modèles UML). Les contraintes seront précisées sur le modèle UML ainsi chargé. Ce modèle UML décrit le Modèle Conceptuel de Données (MCD) sur lequel porteront les contraintes. Le code SQL généré porte sur un Modèle Physique de Données (MPD) produit à partir du modèle UML. OCL2SQL permet aussi de produire automatiquement le MPD à partir du MCD. Des règles de conversion « MCD vers MLD » sont préétablies dans OCL2SQL : une classe du MCD correspond à une table du MPD, une association du MCD peut donner lieu en fonction de ses multiplicités à une table d'association dans le MPD, etc. Plus de détails sur les règles de conversion peuvent être trouvés dans (Demuth & Hussmann, 1999; Demuth, Hussmann & Loecher, 2001).

3.3.2 Principe de l'extension spatiale d'OCL2SQL

Nous avons inclus à OCL2SQL nos extensions spatiales en ajoutant les nouvelles syntaxes OCL proposées et en fournissant dans un premier temps, la génération automatique de triggers et de requêtes pour le SQL spatial supporté par Oracle.

La Figure 15 présente les différents fichiers nécessaires en entrée de l'outil OCL2SQL étendu au spatial : le diagramme de classes UML sous forme d'un fichier XMI, le fichier contenant la ou les contraintes, un fichier contenant des métadonnées sur les données géographiques. Ce dernier fichier spécifie notamment le système de coordonnées spatiales à utiliser pour chaque attribut spatial du modèle UML. L'outil génère en sortie des requêtes et triggers pour le contrôle des contraintes ainsi que différents scripts assurant la création de la base de données spatiales correspondant au diagramme UML. Une structure de données permettant de stocker des régions simples et des régions composites est donc générée dans la base de données (Dubois et al., 2005b). Les requêtes SQL produites peuvent être lancées sur la base de données à des instants choisis par l'administrateur. Ceci permet de détecter les incohérences qui existent à ces instants. C'est cette méthode que nous avons employée lors de notre expérimentation. L'exécution des requêtes au sein des triggers (et donc à chaque mise à jour de la base) peut poser des problèmes de performance pour certains jeux de données. La structure du code généré par OCL2SQL se prête davantage à une exécution sporadique des requêtes.

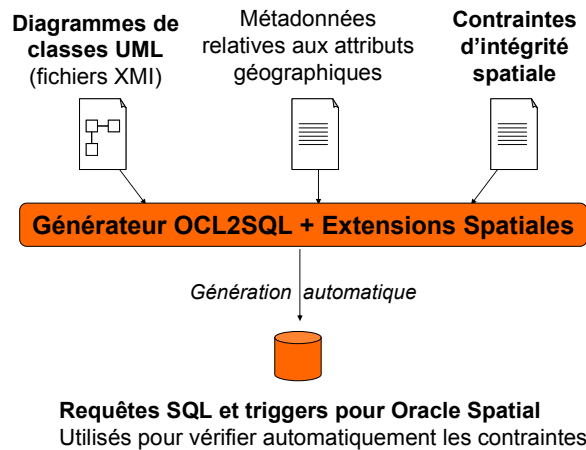


Figure 15. Entrées / sorties du générateur OCL2SQL spatial

L'objectif final est de supporter à terme d'autres SGBD tels que PostGIS par exemple. Pour PostGIS, une génération de code spécifique devra donc être proposée (pour le SQL spatial qu'il supporte).

3.4 Expérimentation sur des données environnementales

Deux expérimentations notables des langages présentés ont été réalisées :

- L'une dans le cadre d'un projet de SI pour le suivi des épandages agricoles. Ce SI a été développé par le Cemagref en lien avec le Ministère de l'Agriculture, le Ministère de l'Environnement, l'ADEME, les Agences de l'Eau. Il s'agit de SIGEMO, le Système Informatisé de Gestion des Epandages de Matières Organiques (Soulignac et al., 2006; Soulignac et al., 2005; Soulignac, Pinet & Gibold, 2004).
- L'autre dans le cadre d'un projet de conception de base de données de relief utilisant un nouveau modèle innovant décrit en détail dans (Saux et al., 2004). Le projet financé par Egide impliquait l'IRENAV, le laboratoire STEM, le Cemagref et le LIMOS.

Dans cette section, nous présentons les tests réalisés avec SIGEMO. Le but est d'illustrer les langages sur un cas réel. Ici, les contraintes sont exprimées en OCL_{9IM}.

3.4.1 Présentation du contexte

SIGEMO permet de gérer les propositions d'épandage agricole de matière organique en France. Il répond à une forte demande d'information émanant de nombreuses institutions. En effet, un encadrement et un suivi rigoureux des retours au sol des matières organiques par les services de l'Etat est indispensable pour assurer de bonnes pratiques d'un point de vue environnemental, économique et sanitaire (Soulignac, 2000).

SIGEMO répond à une nécessité de suivi fiabilisé des pratiques d'épandage de matières organiques en France. Mis en place en 2006, ce système d'information permet à différents acteurs de saisir, transmettre ou analyser des projets d'épandage par Internet. Toutes les informations sont centralisées dans une base de données nationale. Les propositions d'épandage (incluant les parcelles ciblées) sont

tout d'abord entrées dans le SI par des bureaux d'étude pour le compte des producteurs de matières organiques. Ensuite, différents services départementaux et régionaux de l'Etat étudient et contrôlent l'acceptabilité environnementale des propositions de plans d'épandage ainsi saisies dans SIGEMO. De cette étude dépendra l'autorisation d'épandre. Les données de SIGEMO sont également utiles (entre autres) aux collectivités locales et agences de l'eau pour des expertises ou en simple consultation. Avant SIGEMO, les services de l'Etat étudiaient des versions papiers des propositions d'épandage envoyées par les bureaux d'étude (incluant des cartes papiers). L'informatisation fournie par SIGEMO simplifie les tâches des différents acteurs tout en permettant des gains en qualité de traitement et en productivité (Soulignac et al., 2006). De plus l'informatisation des données rend aussi possible des analyses environnementales plus poussées.

Dans la phase de démarrage du projet, différentes données issues de sources externes ont dû être intégrées de façon automatique dans la base nationale - par exemple des propositions de parcelles d'épandage déjà disponibles en version électronique avant SIGEMO. Nous avons appliqué nos propositions de recherche pour modéliser et vérifier les contraintes d'intégrité de SIGEMO, aussi bien sur les données intégrées dans SIGEMO à partir d'autres sources, que sur celles saisies directement par les utilisateurs via Internet après le déploiement de l'application. Dans cette optique, nos travaux ont ainsi visés à offrir, d'une part, des moyens pour spécifier les contraintes alphanumériques et spatiales, et d'autre part, des outils pour tester effectivement ces contraintes sur la base de données afin de contrôler sa qualité.

La base de données de SIGEMO est gérée avec le SGBD Sybase et le SIG MapInfo. Afin de simplifier les tests, nous avons rapatrié toutes les données (de Sybase et de MapInfo) dans une seule base de données sous Oracle 10g, et ce sans modifier la structure générale du schéma. Ainsi, les requêtes SQL pour Oracle générées avec OCL2SQL ont pu être directement testées sur cette version de la base.

Nous avons défini les contraintes d'intégrité puis généré les requêtes SQL correspondantes. Nous avons vérifié les contraintes d'intégrité de façon sporadique sur des jeux de données réels. Les données correspondaient à divers départements. Par exemple, celles concernant la Haute-Loire contenaient 940 parcelles d'épandage. Une vingtaine de contraintes dont une quinzaine de contraintes spatiales ont été modélisées et testées. Cette expérimentation a permis d'identifier les données ne respectant pas certaines règles définies avec les experts métiers.

La Figure 16 présente un extrait simplifié du schéma de la base de données. Les principaux concepts de ce diagramme s'articulent autour des entités suivantes :

- Les dossiers de projet d'épandage (classe *Dossier*). Il s'agit des dossiers de projet d'épandage contenant notamment les propositions des parcelles d'épandage, les produits à épandre, etc. Chaque dossier est réalisé en principe par un bureau d'étude pour le compte d'établissements producteurs de matière, et est soumis par SIGEMO pour évaluation auprès des services de l'Etat. Les informations alphanumériques et géographiques du dossier soumis sont saisies par le Web dans SIGEMO par les bureaux d'étude grâce à une interface SIG spécifique. Après étude, le projet soumis peut être accepté ou non par les services de l'Etat.
- Les parcelles d'épandage (classe *Parcelle*). Il s'agit des parcelles proposées à l'épandage. Il y a dans la grande majorité des cas plusieurs parcelles par dossier.
- Le diagramme présente aussi des classes correspondant aux communes et départements français, ainsi que des classes *ZNIEFF* (Zones Naturelles d'Intérêt Ecologique Faunistique et Floristique) et *Site_Industriel*.

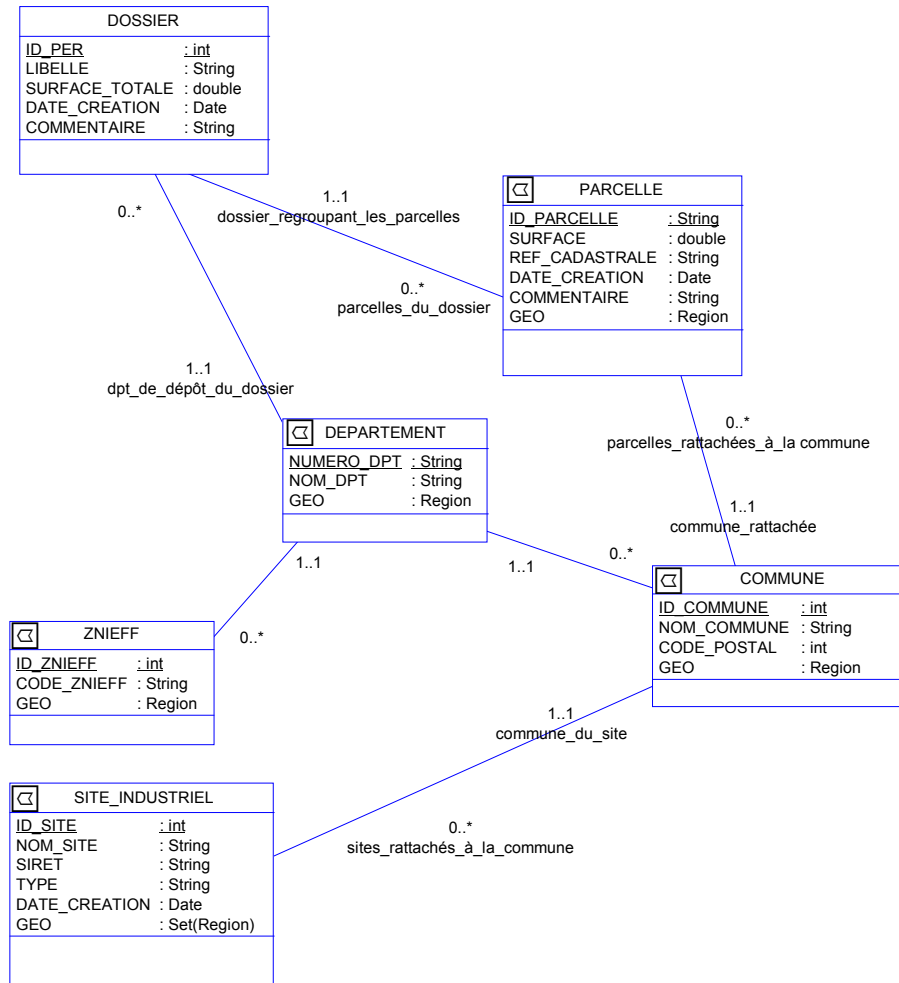


Figure 16. Diagramme de classes de SIGEMO

3.4.2 Exemple de contraintes

Contrainte 1

Dans la base de données, une parcelle d'épandage est associée à une commune principale (correspondant à l'association *commune_rattachée* sur la Figure 16). En pratique, la configuration topologique "disjoint" n'est pas tolérée entre une parcelle d'épandage et sa commune principale. En OCL_{9IM}, cela s'écrit de la façon suivante :

```
context Parcelle inv:
    not ( (self.geo) .disjoint (self.commune_rattachee.geo) )
```

Afin de présenter les correspondances entre cette contrainte OCL_{9IM} et le code SQL généré, les différentes parties de la contrainte ont été numérotées de *p0* à *p5*.

```

      (p0)
context Parcelle inv:
  not ( (self.geo) .disjoint (self.commune_rattachee.geo) )
      (p1)      (p4)      (p2), (p5)      (p3)

```

Le code SQL produit à partir de cette contrainte avec l'extension spatiale de OCL2SQL, est fourni ci-après ; nous avons indiqué les correspondances avec les différentes parties *p0-p5* de la contrainte OCL_{9IM}.

```

(p0) select * from PARCELLE SELF where not
(p1) (NOT (
(p2)   (MDSYS.SDO_RELATE (
(p3)   (select GEO from COMMUNE where ID_COMMUNE in
        (select COMMUNE_RATTACHEE_ID_COMMUNE from PARCELLE
          where ID_PARCELLE = SELF.ID_PARCELLE))
(p4)   , SELF.GEO
(p5)   , 'mask=DISJOINT querytype=WINDOW') = 'TRUE')));

```

On souhaite pouvoir sélectionner les tuples qui violent la contrainte. Ainsi, la requête générée correspond à une négation de la contrainte OCL, ce qui explique le « NOT » en fin de *p0* dans la requête. Le « NOT » qui se trouve en *p1* dans la requête correspond directement au « NOT » spécifié dans la contrainte. La fonction *Mdsys.SDO_Relate* d'Oracle prend ici trois paramètres, à savoir les deux attributs spatiaux et la relation topologique qui doit être testée entre ces deux attributs (ici "Disjoint"). *Commune_Rattachee_ID_Commune* correspond à la clé étrangère présente dans la table *Parcelle*.

Contrainte 2

Les zones ZNIEFF définissent des périmètres d'intérêt écologique. On distingue deux types de ZNIEFF identifiés par un code (*code_znieff*). Aucun épandage ne devrait être proposé dans les zones ZNIEFF de type 1. Autrement dit, les parcelles d'épandage et les zones ZNIEFF de code 1 sont disjointes ou adjacentes.

```

context Parcelle inv:
Znieff.allInstances() -> select(code_znieff = 1) ->
forAll(z | (self.geo) .disjoint (z.geo) or
          (self.geo) .meet      (z.geo) )

```

Cette contrainte dépend d'une condition sur un attribut. En effet, cette contrainte ne s'applique qu'à une partie des instances de la classe *ZNIEFF*, celles dont le code est 1. Ces instances sont retournées grâce à l'opération *select* d'OCL.

Contrainte 3

Un dossier est soumis par SIGEMO auprès des services de l'Etat d'un département pour évaluation. Dans le modèle de la base de données, ce lien entre un dossier et un département est modélisé par l'association *dpt_de_dépôt_du_dossier* (Figure 16). Les parcelles d'épandage d'un dossier sont généralement contenues dans le département qui lui est associé dans la base (c'est-à-dire le

département où il sera évalué). Néanmoins, en pratique il est toléré que le dossier puisse inclure des parcelles qui appartiennent à des départements limitrophes au département d'évaluation. Par exemple, si on considère un dossier instruit par le département 63 (Puy-de-Dôme), les parcelles d'épandage du dossier peuvent alors être dans les départements 63, 03, 42, 43, 15, 19, 23 (voir exemple de la Figure 17). La contrainte est modélisée ainsi :

```
context Dossier inv:
```

```
(1) self.parcelles_du_dossier -> forAll (p |
(2)   Departement.allInstances() -> select (d1 |
(3)     (d1.numero_dpt=self.dpt_de_depot_du_dossier.numero_dpt) or
(4)     (d1.geo) .meet (self.dpt_de_depot_du_dossier.geo))
(5) -> exists(d2| (p.geo) .inside    (d2.geo) or
(6)               (p.geo) .coveredBy (d2.geo) or
(7)               (p.geo) .overlap   (d2.geo)) )
```

La ligne 1 précise que la contrainte doit être satisfaite pour chaque parcelle p d'un dossier. Les lignes 2-4 créent une collection contenant le département d'évaluation du dossier et les départements limitrophes. Les lignes 5-7 indiquent que dans cette collection, il doit exister un département contenant, couvrant ou chevauchant p .

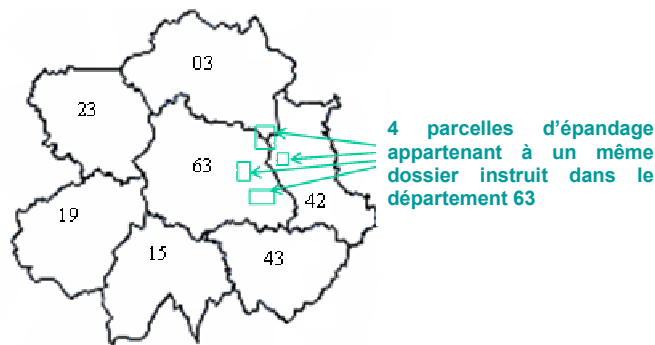


Figure 17. Exemple de parcelles d'épandage se trouvant sur des départements limitrophes du département d'évaluation du dossier (63)

Certains recouvrements des données de SIGEMO avec des données géographiques de type région composite ont aussi été considérés. Des contraintes portant sur ces objets peuvent être définies comme présenté ci-après.

Contrainte 4

Chaque site industriel est représenté par une région composite. Généralement, au moins un bâtiment du site industriel devrait effectivement être spatialement localisé dans sa commune de rattachement. En OCL_{9IM}, cela donne :

```
context Site_Industriel inv:
  self.geo -> exists ( batiment |
    (batiment) .inside (self.commune_du_site.geo) )
```


self.geo retourne la collection de tous les bâtiments d'une instance de la classe *Site_Industriel*.

Contrainte 5.

Afin de présenter différentes possibilités d'OCL_{9IM}, on peut aussi expérimenter une restriction de la contrainte précédente. Faisons l'hypothèse (forte) qu'au moins la moitié des bâtiments d'un site industriel devrait effectivement être spatialement localisée dans sa commune. On peut alors écrire la contrainte suivante.

```
context Site_Industriel inv:
  (self.geo) ->
    select(batiment|batiment .inside (self.commune_du_site.geo))
      -> size() >= (self.geo->size())/2
```

3.5 Conclusion et pistes de recherche à développer

Les objectifs des travaux étaient, d'une part de proposer une méthode expressive pour modéliser les contraintes d'intégrité des bases de données spatiales, et d'autre part, de mettre en place un outil pour les vérifier. Il existait en effet dans ce contexte, un besoin en termes de langage spécialisé, réellement expressif et déclaratif. Le challenge était de permettre l'écriture de contraintes pouvant être complexes, et de générer de façon automatisée les mécanismes de vérification associés. Pour répondre à ce besoin, nous avons proposé, implémenté et expérimenté une adaptation d'OCL pour l'information spatiale, possédant un pouvoir d'expression élevé. En ce qui concerne l'implémentation, nous avons étendu le générateur de code OCL2SQL afin de rendre possible la production de requêtes SQL spatial à partir de contraintes exprimées en OCL. Cet outil a permis de valider nos propositions sur SIGEMO, un système national de gestion des projets d'épandage. Le Tableau 3 présente le positionnement de nos travaux par rapport à l'existant. Notre proposition offre la possibilité de modéliser un large éventail de contraintes d'intégrité spatiale pour contrôler la cohérence sémantique et logique des SI. Avec notre solution, l'écriture de contraintes complexes est réalisée avec l'aide d'informaticiens. Trouver des méthodes pour permettre aux non-informaticiens de modéliser intuitivement des contraintes d'intégrité spatiale complexes nous semble être une tâche particulièrement compliquée que nous pourrions essayer de considérer dans nos travaux futurs. Par contre, l'utilisation conjointe de deux outils pour spécifier l'ensemble des contraintes spatiales d'une même base de données, pourrait être envisagée. Par exemple les contraintes d'un niveau de complexité moyen pourraient être modélisées avec les outils orientés utilisateurs finaux (i.e. les interfaces visuelles de définition de contraintes spatiales) tels que ceux présentés en Section 2.2.1. Les contraintes plus complexes seront écrites en OCL étendu au spatial, avec OCL2SQL. Les outils orientés utilisateurs finaux pourraient aussi être étendus afin de générer des contraintes en OCL spatial. Ceci permettrait d'avoir en OCL, l'ensemble des contraintes d'une base de données. Un schéma d'utilisation conjointe de ces outils est présenté à la Figure 18.

| | Type d'utilisateurs | Niveau d'abstraction | Expressivité | Spécialisation pour l'expression de contraintes |
|---|---|----------------------|--------------|---|
| Interfaces visuelles de définition de contraintes spatiales | Non-informaticiens | Modèle Conceptuel | + | ++ |
| Langage de contraintes OCL étendu au spatial | Principalement informaticiens ; utilisation par des non-informaticiens envisageable pour la modélisation de contraintes simples | Modèle Conceptuel | ++ | ++ |
| Langages SQL spatial | Informaticiens | Modèle Physique | ++ | + |

Tableau 3. Positionnement de notre proposition par rapport à l'existant

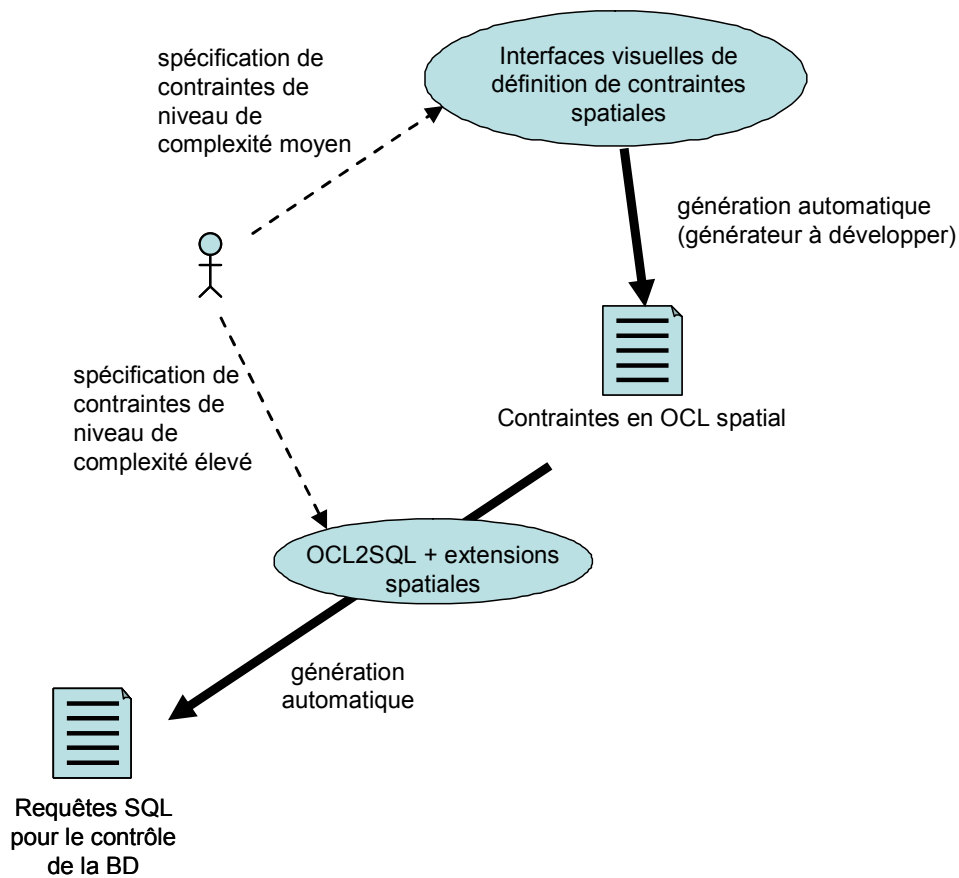


Figure 18. Couplage d'OCL2SQL avec les outils de type interface visuelle de définition de contraintes spatiales

Suite à notre expérimentation, la démarche méthodologique générale de spécification des contraintes la plus cohérente nous semble être la suivante (dans les grandes lignes). Lors de la constitution du Modèle Conceptuel de Données (MCD), les contraintes d'intégrité peuvent être écrites en langage naturel ou avec OCL étendu au spatial. Puis elles pourront être raffinées jusqu'à la production du Modèle Physique de Données. A cette étape, toutes les contraintes d'intégrité doivent être exprimées en OCL spatial, ce qui permettra la génération de code SQL avec OCL2SQL. L'écriture des contraintes avec OCL se fait alors à partir de la dernière version du MCD. Ceci nécessite d'avoir toujours une version du MCD à jour, c'est-à-dire reflétant exactement la structure de la base de données.

Dans les travaux présentés dans ce mémoire, nous soulignons qu'il est par nature plus facile d'exprimer les contraintes sur un MCD. Néanmoins, nous pensons que dans certains cas, une modélisation des contraintes d'intégrité en OCL au niveau du Modèle Logique de Données (MLD) pourrait présenter un intérêt ; par exemple dans le cas de MLD qui évoluent très vite sans qu'une réactualisation du MCD soit réalisée. Dans ce cas, il faudrait étudier comment des contraintes OCL peuvent être spécifiées au niveau d'un MLD.

Pour nos travaux nous avons aménagé la fonctionnalité de OCL2SQL qui permet de générer des requêtes SQL de contrôle. Ces requêtes sont exécutées sur les bases de données à des instants choisis par les administrateurs et détectent donc les incohérences qui existent à ces instants. Il est possible de générer aussi le code de triggers qui seraient déclenchés automatiquement à chaque mise à jour des bases. Des problèmes de performances peuvent alors se poser dans le cas de grandes bases de données. L'optimisation de ces performances implique à notre avis un travail de recherche assez complexe. Il s'agit d'une piste de travail très riche.

Au final, les résultats obtenus nous conduisent à penser que les pistes d'utilisation d'OCL pour les contraintes d'intégrité spatiale devraient être poursuivies, par exemple en étendant les fonctionnalités du langage.

Une prochaine étape pourrait être d'élargir le champ des recherches à d'autres types d'objets : relations topologiques entre points, entre lignes, entre objets de types différents (Egenhofer & Herring, 1992), entre collections de régions simples pouvant être adjacentes et/ou intégrant des trous (régions complexes) (Behr & Schneider, 2001; Egenhofer, Clementini & Di Felice, 1994; Nguyen, Parent & Spaccapietra, 1997), entre collections d'objets de types différents (Zhong et al., 2004), etc. Dans un deuxième temps, d'autres relations spatiales, comme par exemple les relations métriques, mériteraient d'être considérées. Dans SIGEMO, une distance suffisante doit être respectée entre une parcelle acceptée pour l'épandage et certaines zones hydrologiques avoisinantes. D'une manière plus générale, nous pourrions proposer assez facilement l'intégration dans OCL de différentes opérations de SQL spatial telles que *Union*, *Envelope*, *Centroid*, *Area*, etc. (OpenGIS, 1999). Par exemple, l'utilisation d'une fonction d'union spatiale simplifierait l'écriture de la contrainte 3 de la Section 3.4.2. En effet, il serait alors possible de créer une région agrégeant le département d'évaluation avec tous ses départements limitrophes. Les relations spatiales avec les parcelles d'épandage ne seraient ensuite contrôlées que sur cette région agrégée. Enfin, la représentation des contraintes d'intégrité spatio-temporelle avec OCL est une autre perspective.

Nous démarrons actuellement avec Catherine Roussey, Maître de Conférences en délégation dans notre équipe depuis septembre 2009, des recherches sur un outil destiné à contrôler si un ensemble de contraintes d'intégrité topologique est cohérent. Ces recherches se basent sur une description des relations topologiques en Logique de Description (Katz & Cuenca Grau, 2005; Pinet et al., 2009b; Pinet et al., 2006; Roussey et al., 2010).

Suite aux travaux résumés dans le présent chapitre, nous avons décidé de mener différentes investigations sur l'utilisation d'OCL dans des bases de données contenant des régions incertaines. L'objectif était d'exprimer des contraintes sur des régions ayant des frontières mal délimitées. Cette incertitude des frontières est caractéristique de beaucoup d'objets environnementaux (peuplements

forestiers, lacs, zones de pollution, etc.). Dans le cadre de l'application de nos travaux au domaine environnemental, il nous a donc semblé essentiel de privilégier cette piste qui va être décrite dans le chapitre suivant.

4 Modélisation de contraintes d'intégrité topologique impliquant des régions vagues

4.1 Introduction

Les extensions spatiales d'OCL du chapitre précédent ne sont pas conçues pour modéliser des contraintes topologiques impliquant des objets aux limites incertaines. Pourtant, les limites de nombreux objets environnementaux sont mal connues ou imprécises (Bejaoui et al., 2008a; Burrough & Frank, 1996; Clementini & Di Felice, 1997; Cohn & Gotts, 1996; Dilo, 2006; Reis, Egenhofer & Matos, 2006; Tang, 2004). Par exemple, (Miralles, 2006) donnent des illustrations d'objets incertains que l'on peut trouver dans les SI environnementaux. Il est donné l'exemple de contours incertains de parcelles de blé estimés à partir d'images satellitaires ; le contour n'est pas clairement délimité car, avec les techniques utilisées de traitement d'images, on ne peut être certain de la présence de blé en certains pixels de l'image.

Dans ce chapitre, nous présentons les recherches que nous avons menées pour :

- améliorer les méthodes existantes de représentation des objets spatiaux incertains,
- déterminer un ensemble de relations topologiques entre ces objets,
- intégrer ces relations topologiques dans OCL.

Ces recherches ont été réalisées dans le cadre de la thèse de doctorat de Lotfi Bejaoui (Bejaoui, 2009) préparée en cotutelle (France-Canada). Il s'agit d'une collaboration entre le Cemagref, l'équipe d'Yvan Bédard de l'Université Laval et le LIMOS.

Dans ce chapitre, nous nous focaliserons sur les régions spatiales incertaines. Dans la Section 4.2, nous présentons un état de l'art ainsi que notre proposition de modèle de régions et de relations topologiques associées. Des versions plus détaillées des états de l'art et de nos propositions peuvent être trouvées dans (Bejaoui et al., 2008a). Dans la section 4.3, nous introduisons notre extension d'OCL spatial. La Section 4.4 présente une expérimentation sur les données de SIGEMO. La Section 4.5 présente les conclusions et les perspectives de notre travail.

4.2 Les régions incertaines et leurs relations topologiques

4.2.1 Modèles antérieurs à nos propositions

La théorie dite du jaune d'œuf (Cohn & Gotts, 1996) introduit le concept de régions spatiales ayant une frontière incertaine. En fait, ce type de région spatiale appelée « vague » se compose de deux régions simples : la région intérieure (le « jaune ») entouré d'une région externe (le « blanc »). La région du jaune est la partie certaine de l'objet. La région du blanc est la partie incertaine ; elle correspond à la frontière incertaine de l'objet (voir Figure 19). Les deux régions sont adjacentes et leur union compose « l'œuf » c'est-à-dire l'objet vague. Dans (Cohn & Gotts, 1996), le modèle proposé distingue 46 relations topologiques entre deux régions vagues de type « jaune d'œuf ». (Clementini & Di Felice, 1997) ont proposé un modèle sémantiquement équivalent à celui du jaune d'œuf, mais avec une définition légèrement différente. Dans ce modèle, une région vague A est composé de deux régions simples A1 et A2 ; A1 est incluse dans A2. La différence entre A2 et A1 donne la partie incertaine. Avec leur modèle, (Clementini & Di Felice, 1997) distingue 44 relations topologiques entre deux

régions vagues. Plus récemment, (Xinming, 2004) a étendu cette approche pour pouvoir proposer un modèle distinguant plus de relations : 152 relations topologiques ont été proposées.

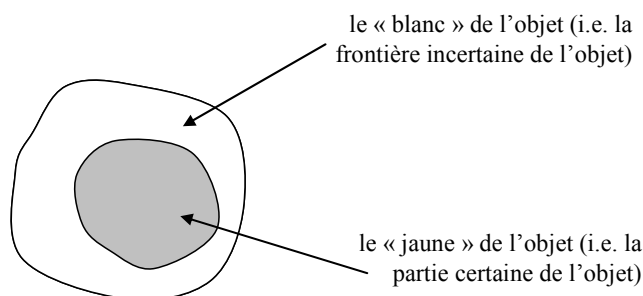


Figure 19. Une région spatiale vague selon le modèle du jaune d'œuf

La théorie des sous-ensembles flous (Zadeh, 1965) est également utilisée pour représenter des objets spatiaux incertains (Dilo, 2006; Schneider, 2001; Schneider, 2008; Shu et al., 2003; Yazici & Akkaya, 2000; Yazici, Zhu & Sun, 2001). Une région spatiale floue est modélisée par un sous-ensemble flou. Dans ce mode de représentation, une fonction d'appartenance est associée à chaque objet spatial incertain. Elle retourne pour tout point de l'espace, le degré d'appartenance à l'objet. Si un point de l'espace est associé un degré égal à 1, alors il appartient à l'objet spatial. Si un point a un degré de 0 alors il n'appartient pas à l'objet. Si un point est associé à un degré d compris dans l'intervalle $]0,1[$, alors le point appartient à l'objet à hauteur de ce degré d . La Figure 20 montre un exemple de représentation visuelle de deux régions spatiales floues. Plus la couleur est sombre, plus le degré se rapproche de la valeur 1 (i.e. de la certitude). (Dilo, 2006) propose un modèle de relations topologiques pour cette approche. Entre deux régions spatiales floues (i.e. deux sous-ensembles flous), il définit six relations topologiques. Une relation topologique est modélisée par une fonction prenant en paramètre les deux sous-ensembles flous. La fonction retourne un degré de vérité compris entre 0 et 1 (par exemple $\text{Overlap}(\text{Objet1}, \text{Objet2}) = 0.2$). Ce degré indique à quel niveau la relation est susceptible d'exister entre les deux objets. Une approche ayant un principe similaire est proposée dans (Schneider, 2001; Schneider, 2008).

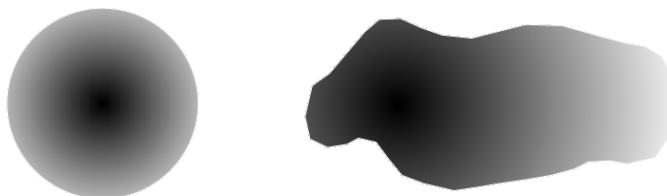


Figure 20. Deux régions spatiales floues (exemple issu de (Schneider, 2008))

Concernant les formalismes ER et OO pour l'information spatiale, les travaux de (Miralles, 2006; Shu et al., 2003; Yazici & Akkaya, 2000; Yazici, Zhu & Sun, 2001) proposent des pictogrammes spécifiques pour indiquer qu'un attribut d'une classe va représenter un objet incertain.

4.2.2 Un nouveau modèle de régions incertaines

Les modèles d'objets flous nécessitent de posséder beaucoup d'informations sur les objets observés. En effet, il faut pouvoir estimer un degré de certitude en tout point des objets (Dilo, 2006; Schneider, 2001; Schneider, 2008). Cette contrainte peut rendre assez difficile la mise en pratique de ces modèles.

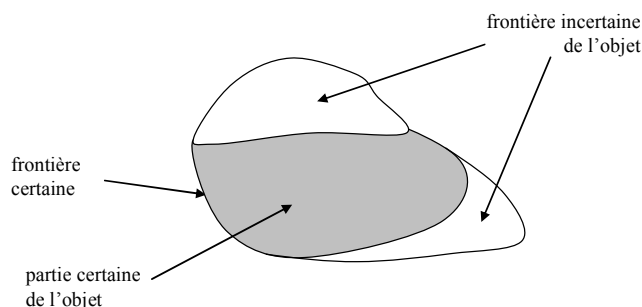


Figure 21. Une région spatiale vague à la frontière partiellement vague

Nous nous sommes donc plutôt intéressés aux modèles de régions vagues proposés par (Clementini & Di Felice, 1997; Cohn & Gotts, 1996; Xinming, 2004) (Figure 19). Ces modèles, bien que moins « expressifs » que ceux basés sur les sous-ensembles flous, ne nécessitent pas de méthodes complexes d'estimation de degré de certitude. Nous avons étendu le modèle du jaune d'œuf afin de pouvoir modéliser certains objets courants dans le domaine de l'environnement.

En effet, les propositions de (Clementini & Di Felice, 1997; Cohn & Gotts, 1996; Xinming, 2004) considèrent que la frontière des régions vagues est toujours totalement incertaine. Ces modèles ne permettent pas de représenter les régions ayant une frontière partiellement incertaine (c'est-à-dire les régions avec une frontière ayant à la fois des parties certaines et des parties incertaines). La Figure 21 donne un exemple. Nous avons donc étendu ces modèles afin d'y ajouter ce type d'objet. Notre nouveau modèle appelé « Qualitative Min-Max » (*QMM*) et détaillé dans (Bejaoui et al., 2008a; Bejaoui et al., 2008b) permet de représenter, à la fois, les régions ayant une frontière totalement incertaine (Clementini & Di Felice, 1997; Cohn & Gotts, 1996; Xinming, 2004) et les régions ayant une frontière partiellement incertaine. Dans notre modèle, une région vague est composée de deux régions simples A_1 et A_2 . A_1 peut être soit incluse dans A_2 (Figure 19) soit couverte par A_2 (Figure 21). Nous appelons A_1 (resp. A_2), l'extension minimale (resp. maximale) de la région vague. Dans la suite du chapitre, nous noterons $\min(A)$, l'extension minimale, et $\max(A)$, l'extension maximale. La différence entre $\max(A)$ et $\min(A)$ donne la partie incertaine de A .

Un nouveau modèle de relation topologique, adapté au type de régions vagues proposé, a été élaboré (Bejaoui et al., 2008a; Bejaoui et al., 2008b). Dans ce modèle, chaque relation topologique est représentée par une matrice 2×2 :

| | | |
|-----------|------------------------|------------------------|
| | $\min(B)$ | $\max(B)$ |
| $\min(A)$ | $SR(\min(A), \min(B))$ | $SR(\min(A), \max(B))$ |
| $\max(A)$ | $SR(\max(A), \min(B))$ | $SR(\max(A), \max(B))$ |

Dans les cellules de la matrice, $SR(O1, O2)$ indique la relation topologique entre $O1$ et $O2$. Dans notre modèle, $SR(O1, O2)$ est égal à l'une des relations topologiques d'Egenhofer décrites à la Figure 11 : *disjoint*, *contains*, *inside*, *equal*, *meet*, *covers*, *covered by*, *overlap*. Nous avons démontré dans (Bejaoui et al., 2008a) que ce modèle permettait de distinguer 242 relations topologiques (représentées par 242 matrices) entre deux régions vagues ayant des frontières partiellement ou totalement incertaines. Cinq exemples de relations topologiques sont décrits à la Figure 22. Dans la suite de cette sous-section, nous utiliserons le terme « sous-relation » pour parler des relations $SR(O1, O2)$ présentes dans les cellules des matrices, et nous emploierons le terme « relation » pour parler de l'une des 242 relations.

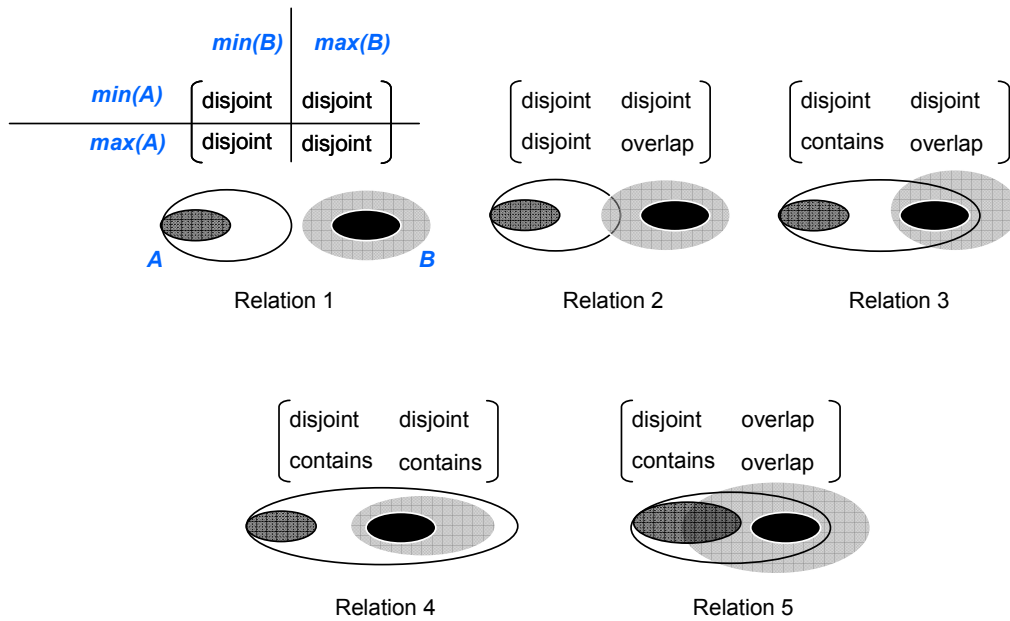


Figure 22. Cinq exemples de relations entre deux régions vagues selon le modèle *QMM* de (Bejaoui et al., 2008a)

Dans un but de faciliter leur utilisation au sein de langage informatique comme OCL, nous avons proposé de regrouper les 242 relations (Bejaoui et al., 2008a). Un groupe de relations est identifié par un couple (*adverbe*, *Nom de sous-relation*). *Nom de sous-relation* peut être le nom d'une des 8 relations d'Egenhofer décrites à la Figure 11, et *adverbe* $\in \{weakly, fairly, strongly \text{ et } completely\}$:

- une relation R appartient au groupe (*weakly*, *Nom de sous-relation*) si *Nom de sous-relation* apparaît une et une seule fois dans la matrice représentant R .
- une relation R appartient au groupe (*fairly*, *Nom de sous-relation*) si *Nom de sous-relation* apparaît exactement deux fois dans la matrice représentant R .
- une relation R appartient au groupe (*strongly*, *Nom de sous-relation*) si *Nom de sous-relation* apparaît exactement trois fois dans la matrice représentant R .
- une relation R appartient au groupe (*completely*, *Nom de sous-relation*) si *Nom de sous-relation* apparaît exactement quatre fois dans la matrice représentant R .

Par exemple, dans la Figure 22, la relation 1 se trouve dans le groupe (*completely*, *Disjoint*) car *Disjoint* apparaît quatre fois dans la matrice. La relation 2 est dans le groupe (*strongly*, *Disjoint*), les

relations 3 et 4 sont dans (*fairly, Disjoint*), et la relation 5 dans (*weakly, Disjoint*). A noter qu'une relation peut se trouver dans plusieurs groupes. Par exemple, la relation 2 de la Figure 22 appartient à la fois au groupe (*strongly, Disjoint*) et au groupe (*weakly, Overlap*), la relation 3 est simultanément dans (*fairly, Disjoint*), (*weakly, Contains*) et (*weakly, Overlap*).

4.3 Extension d'OCL pour les régions spatiales incertaines

Nous avons proposé dans (Bejaoui et al., 2008b, 2010) une nouvelle extension d'OCL nommée OCL_{OVS} intégrant les relations topologiques du modèle *QMM* présenté à la sous-section précédente.

Définition 3.1. Région vague (avec frontière totalement ou partiellement incertaine).

Une région vague (simple) A est composée d'une région simple $\min(A)$ incluse dans ou couverte par une région simple $\max(A)$.

□

Le type *Region_Vague* est inclus dans le langage OCL ainsi que de nouvelles opérations pour vérifier les relations topologiques entre les objets de ce type.

Définition 3.2. Opérations spatiales de OCL_{OVS} .

La syntaxe générale de ces opérations spatiales est décrite ci-dessous.

```
(vagueGeoA).adverbe_opérationTopo(vagueGeoA) : Boolean
```

où *adverbe* $\in \{weakly, fairly, strongly, completely\}$ et *opérationTopo* $\in \{Inside, Contains, Covers, CoveredBy, Disjoint, Equal, Overlap, Meet\}$; *vagueGeoA*, *vagueGeoB* sont de type *Region_Vague* (avec frontière totalement ou partiellement incertaine).

vagueGeoA et *vagueGeoB* sont les paramètres de l'opération. Chaque opération retourne vrai ou faux (un booléen) suivant que la relation topologique entre *vagueGeoA* et *vagueGeoB* appartient au groupe de relations (*adverbe, opérationTopo*) ou non.

□

Nous illustrons ci-dessous notre langage OCL_{OVS} par quelques exemples basiques.

Contrainte 1

La contrainte suivante indique que les régions vagues doivent être complètement disjointes (cf. relation 1 de la Figure 22)

```
context Class1 inv:
  Class1.allInstances()->forall(o1, o2 |
    (o1.vagueGeo).completely_Disjoint(o2.vagueGeo) )
```

Contrainte 2

La contrainte indique que la relation entre les régions vagues doit être incluse dans le groupe de relations (*fairly*, *Disjoint*). Ce groupe inclut, entre autre, les relations 3 et 4 présentées à la Figure 22.

```
context Class1 inv:
  Class1.allInstances()->forall(o1, o2 |
    (o1.vagueGeo) .fairly_Disjoint (o2.vagueGeo) )
```

Contrainte 3

En OCL_{ADV}, pour pouvoir exprimer exactement l'une des 242 relations du modèle *QMM*, il faut préciser de manière exhaustive tous les groupes auxquels la relation appartient. Par exemple, la contrainte 3 de la Figure 22 peut être modélisée de la façon suivante.

```
context Class1 inv:
  Class1.allInstances()->forall(o1, o2 |
    (o1.vagueGeo) .fairly_Disjoint (o2.vagueGeo) ) and
    (o1.vagueGeo) .weakly_Contains (o2.vagueGeo) ) and
    (o1.vagueGeo) .weakly_Overlap (o2.vagueGeo) )
```

Contrainte 4

Les contraintes suivantes indiquent que la relation entre les régions vagues doit être incluse dans au moins un des groupes suivants : (*fairly*, *Disjoint*), (*weakly*, *Contains*), (*weakly*, *Overlap*)

```
context Class1 inv:
  Class1.allInstances()->forall(o1, o2 |
    (o1.vagueGeo) .fairly_Disjoint (o2.vagueGeo) ) or
    (o1.vagueGeo) .weakly_Contains (o2.vagueGeo) ) or
    (o1.vagueGeo) .weakly_Overlap (o2.vagueGeo) )
```

A noter qu'il était difficile d'intégrer 242 opérations dans OCL (chaque opération vérifiant une relation topologique en particulier). Avec ce scénario, l'utilisateur du langage doit en effet, toujours avoir à l'esprit la sémantique des 242 opérations. Par contre, l'expression par le biais d'adverbes, telle qu'illustrer dans les quatre contraintes précédentes, offre une solution plus intuitive. Dans le langage proposé, l'adverbe permet d'indiquer un « degré » souhaité d'intersection, de disjonction, de contenance, de couverture ou d'égalité entre deux régions vagues.

Il est aussi facilement possible d'intégrer dans OCL une fonction (appelée par exemple *relationQMM*) qui permettrait de spécifier en paramètre, les 4 sous-relations de la matrice (en plus des deux objets vagues). Dans ce cas, (*o1.vagueGeo*) .*relationQMM* ("*disjoint*", "*overlap*", "*contains*", "*overlap*", *o2.vagueGeo*) retournera vrai si la relation 5 de la Figure 22 est vérifiée entre l'objet *o1.vagueGeo* et l'objet *o2.vagueGeo* (faux dans le cas contraire).

Comme pour nos extensions d'OCL décrites précédemment (voir Section 3.3.2), nous avons intégré OCL_{OVS} dans OCL2SQL. Ainsi, lorsque du code est généré avec OCL2SQL à partir d'une contrainte en OCL_{OVS}, la requête SQL produite utilise des fonctions PL-SQL que nous avons spécifiquement développées pour tester une relation topologique entre régions vagues (*fairly_Disjoint*, *weakly_Contains*, etc.). Une structure adaptée est aussi utilisée pour stocker les régions vagues. Tous les détails de cette implémentation peuvent être trouvés dans (Bejaoui et al., 2008b, 2010).

4.4 Expérimentation sur des données environnementales

Pour illustrer l'utilisation de notre extension d'OCL spatial pour les régions vagues, nous présentons les expérimentations réalisées dans le cadre de SIGEMO (système décrit dans la Section 3.4.1).

Dans SIGEMO, les parcelles d'épandages sont déclarées grâce à un outil Web. Dans la pratique, une valeur numérique indiquant la surface (en ha) de la parcelle est déclarée, puis les contours de cette parcelle sont dessinés à la souris à l'aide de fonds de plan et de l'interface SIG-Web de SIGEMO. Ces deux informations sont stockées dans la base de données nationale de SIGEMO. La surface en ha indiquée est fiable car elle est déterminée par expertise du terrain. L'objectif du dessin des parcelles est d'avoir une indication de la zone impliquée par l'épandage. La surface en ha de la parcelle d'épandage est généralement différente de la surface de la version dessinée à l'écran. Cette différence provient du mode de saisie de la parcelle (à la souris). En pratique, cette différence ne pose pas réellement de problème, la précision du dessin étant suffisante pour avoir une idée du lieu d'épandage.

Dans le cadre de la thèse de Lotfi Bejaoui (Bejaoui, 2009), à titre expérimental, nous avons étudié les moyens de corriger la géométrie des parcelles dessinées afin qu'elles correspondent mieux à la surface déclarée en ha (et donc à la géométrie réelle de la parcelle). Nous avons utilisé notre modèle de région vague proposé pour représenter les parcelles. L'extension minimale de la région correspond à une zone d'épandage certaine et les frontières larges correspondent à des zones d'incertitude i.e. des zones qui sont peut être des lieux d'épandage. A partir des parcelles d'épandage dessinées du jeu de données que nous avons, nous avons créé des « parcelles vagues » i.e. des régions vagues. Le principe est illustré à la Figure 23. Différentes contraintes ont été spécifiées avec OCL_{OVs}. Elles concernaient la modélisation des configurations topologiques considérées comme acceptables entre les différents objets spatiaux e.g. entre les parcelles d'épandage et entre certaines sous-parties des parcelles qui doivent recevoir des traitements spécifiques. Les détails de cette expérimentation sont décrits dans (Bejaoui, 2009; Bejaoui et al., 2008b, 2010).

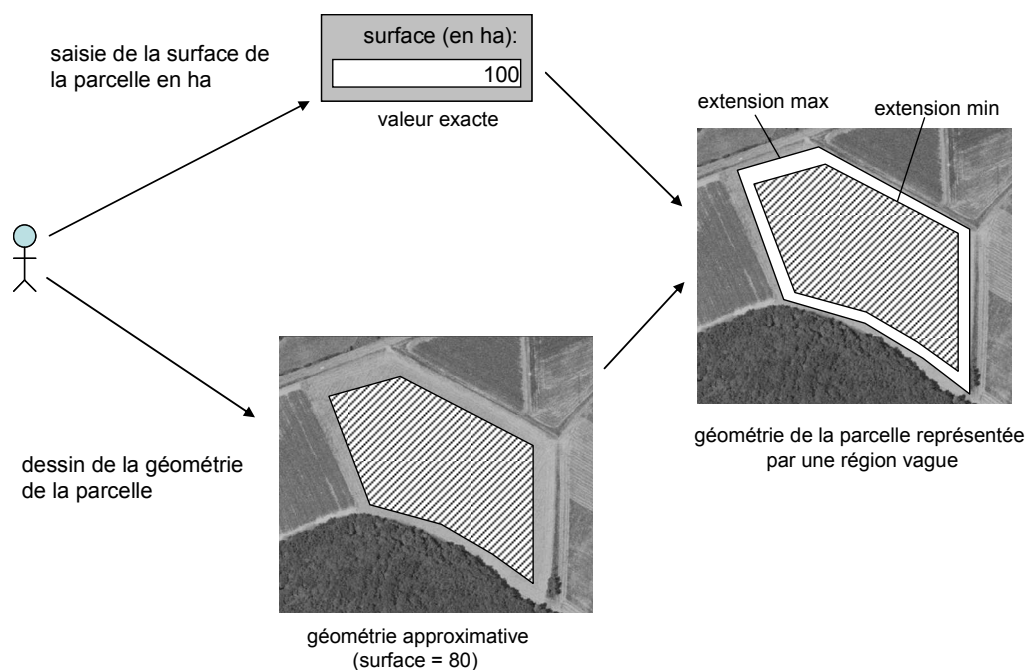


Figure 23. Correction de la géométrie des parcelles

4.5 Conclusion et pistes de recherche à développer

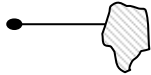
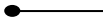
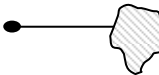
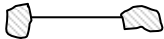

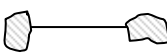








| Lignes vagues | Extensions | |
|--|--------------------|--|
|  | Extension minimale |  |
| | Extension maximale |  |
|  | Extension minimale |  |
| | Extension maximale |  |
|  | Extension minimale |  |
| | Extension maximale |  |
|  | Extension minimale |  |
| | Extension maximale |  |
| <p>● : extrémité connue (certaine) de la ligne</p> <p>— : intérieur connu (certain) de la ligne</p> <p> : partie incertaine de l'intérieur de la ligne (i.e. périmètre où se trouve tout ou partie de l'intérieur « réel » de la ligne)</p> <p> : extrémité incertaine (i.e. périmètre dans lequel se trouve l'extrémité « réelle » de la ligne)</p> | | |

Tableau 4. Exemple de ligne incertaine selon le modèle *QMM* (Bejaoui et al., 2009)

La modélisation des objets spatiaux incertains est un domaine de recherche important. Se posera aussi le problème de la représentation physique dans les bases de données en vue d’optimiser leur stockage et leur traitement.

Ce chapitre a décrit notre contribution pour représenter les régions vagues (avec frontières totalement ou partiellement incertaines), leurs relations topologiques ainsi que les contraintes d'intégrité les impliquant. Nous avons démontré qu'il existe un assez grand nombre de relations topologiques entre régions vagues (242). Se posait le problème d'exprimer facilement ces relations en OCL. L'approche adverbiale que nous avons introduite permet de modéliser ces relations de manière intuitive. Dans le langage proposé, l'adverbe permet d'indiquer un « degré » souhaité d'intersection, de disjonction, de contenance, de couverture ou d'égalité entre deux régions vagues. Une approche adverbiale (OCL_{ADV}) basée sur (Claramunt, 2000) avait déjà été proposée (cf. Chapitre 3) pour exprimer des contraintes entre régions composites. Nous avons ajouté nos propositions dans OCL2SQL. Nous avons donc actuellement un outil de type générateur de code supportant notre « OCL spatial » (pour les régions vagues et pour les régions certaines).

Bien que nos travaux aient principalement portés sur les objets de type région, nous nous sommes aussi intéressés aux objets linéaires et ponctuels vagues. Le modèle *QMM* a été adapté pour les lignes et les points vagues (Bejaoui et al., 2009; Bejaoui et al., 2008a). Le Tableau 4 présente quelques exemples de lignes vagues. Ce modèle pourrait être utilisé pour représenter des réseaux (e.g. hydrologiques) ou des itinéraires, comprenant des parties non connues. Dans (Bejaoui et al., 2009), des relations topologiques ont été définies pour les lignes vagues. Actuellement ces modèles ne sont pas implémentés dans OCL2SQL.

L'approche « extension minimale-maximale » et l'approche adverbiale introduites dans ce chapitre pourraient être adaptées à de nouveaux type d'objets vagues (objets composites, hétérogènes, régions vagues ayant des parties certaines non connexes, etc.). Concernant le stockage de ces objets dans une base de données, il faudrait introduire des structures de stockage performantes pour ces nouveaux types d'objets vagues. A l'heure d'aujourd'hui, les SGBD ne proposent pas de type de données « objets spatiaux vagues » prédéfini.

Le modèle de régions vagues proposé ici est particulièrement adapté pour représenter des données intégrées issues de plusieurs sources (dans le cadre d'une base ou d'un entrepôt de données). Il permet de décrire une « vue intégrée consensuelle » d'un objet. La Figure 24 illustre l'utilisation de régions vagues lors d'une intégration. Dans cet exemple, trois sources de données proposent chacune une représentation différente d'un lac. Cette différence peut provenir de l'utilisation de moyens différents pour déterminer le contour du lac. En l'absence d'information sur la confiance à accorder à chacune des sources, une solution d'intégration consiste à produire une région vague à partir des objets des trois sources. L'extension minimale correspond à la partie certaine de l'objet, c'est-à-dire la partie faisant l'objet d'un consensus. Il s'agit de l'intersection spatiale des différentes représentations de l'objet dans les sources. L'extension maximale est l'union de toutes les représentations. (Bejaoui, 2009) a proposé une étude visant à déterminer les contraintes d'intégrité de la base de données intégrée à partir des contraintes d'intégrité des sources. L'objectif est de pouvoir intégrer non seulement les données des sources mais aussi leurs contraintes.

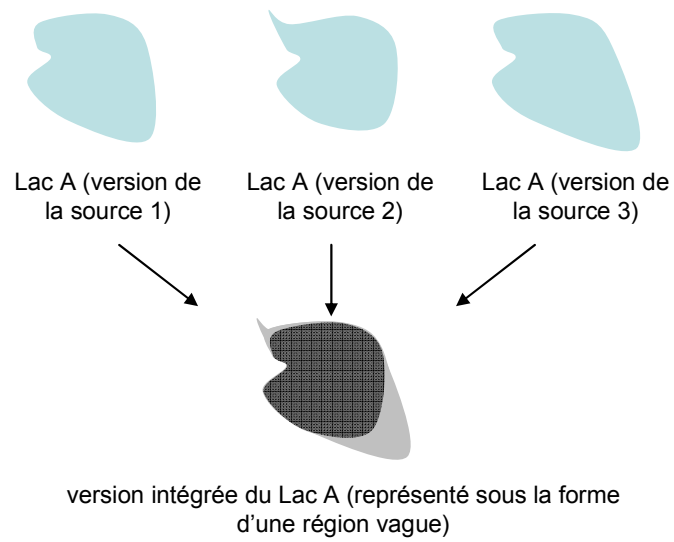


Figure 24. Intégration de trois sources de données stockant chacune le même objet

5 Contraintes d'intégrité dans les entrepôts de données

5.1 Introduction

Avec l'évolution des Nouvelles Technologies de l'Information et de la Communication, nous assistons à une multiplication des sources de données agricoles et environnementales géoréférencées :

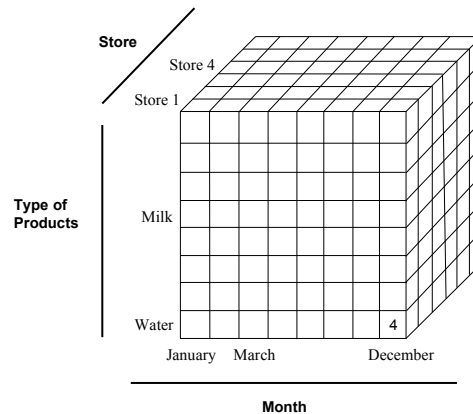
- certaines données sont acquises automatiquement par l'intermédiaire de capteurs ou de dispositifs de télédétection - par exemple, à partir de dispositifs intégrés à des épandeurs pour suivre le travail dans les champs, à partir de réseaux de capteurs, d'images satellites,
- d'autres sont saisies manuellement par un utilisateur grâce à des programmes informatiques spécialisés - par exemple à partir d'une application informatique d'enregistrement des pratiques.

Ces informations nombreuses nécessitent des moyens performants de stockage mais aussi d'intégration et d'analyse. Le concept d'entrepôt de données représente le dispositif moderne le plus approprié pour mobiliser ces moyens. Un entrepôt est un nouveau type de base de données servant à intégrer, accumuler et analyser des données en provenance de diverses sources (Cali et al., 2003). Les informations de différentes bases de données sont chargées dans un entrepôt afin d'y être analysées ensemble. En fonction des besoins, ce chargement peut se faire toutes les semaines, tous les mois, tous les ans, ou encore moins fréquemment. Les outils de type ETL (Extraction, Transformation, Loading) ont pour but d'aider l'intégration des différentes sources dans l'entrepôt (Trujillo & Luján-Mora, 2003).

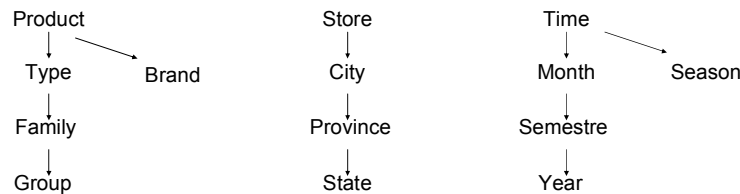
Les données d'un entrepôt sont généralement organisées sous une forme dite multidimensionnelle pour faciliter le calcul d'indicateurs en croisant différents critères. Les indicateurs sont des informations agrégées obtenues par des fonctions d'agrégat telles que la somme, la moyenne, la variance, etc. Les principaux résultats des calculs sont souvent stockés directement dans l'entrepôt. On peut croiser les données d'un entrepôt afin de faire émerger des relations causales inconnues jusqu'alors. Pour cela, les utilisateurs peuvent visualiser les données de l'entrepôt à l'aide d'outils de type OLAP (On-line Analytical Processing). Des outils SOLAP (Spatial On-line Analytical Processing) existent aussi pour l'analyse et la visualisation des données géographiques (Bedard, 2009; Bimonte, 2007).

L'utilisation des entrepôts de données est importante dans un contexte de prise de décision. Cette technologie a été grandement utilisée dans le secteur de la grande distribution. D'autres domaines d'applications sont possibles. Un entrepôt contenant des informations économiques, urbaines et environnementales aidera à trouver le meilleur lieu pour l'installation d'une nouvelle infrastructure. L'analyse et le recoupement des informations de différentes sources au sein d'un entrepôt de données peuvent aussi se faire dans un but d'étude épidémiologique – voir par exemple les entrepôts de (World Health Organization, 2009) et (Bernier et al., 2008). Les entrepôts de données environnementales sont encore assez rares. Ils présenteraient pourtant un fort intérêt. On trouve quelques exemples d'entrepôts environnementaux et agricoles dans (Bernier et al., 2008; Nilakanta, Scheibe & Rai, 2008; Schulze, Spilke & Lehner, 2007). De notre côté, nous travaillons sur la conception et la mise en place d'un entrepôt intégrant plusieurs sources de données géoréférencées relatives à l'utilisation des pesticides en agriculture. Ces travaux se font dans le cadre du projet « SIE Pesticide » financé par le Cemagref. Ce projet implique les centres du Cemagref de Montpellier, Clermont-Ferrand, Lyon, et Bordeaux. L'outil SOLAP utilisé est celui de (Bedard, 2009). Nous travaillons aussi sur une structure générique d'entrepôts de données de résultats de simulations environnementales (Mahboubi et al., 2010). Nous allons utiliser ce modèle avec le Laboratoire d'Ingénierie des Systèmes Complexes du Cemagref, dans

le cadre du projet européen « Prototypical Policy Impacts on Multifunctional Activities in Rural Municipalities » (PRIMA). L'objectif du projet est de concevoir des modèles de simulation pour prévoir les évolutions des différentes catégories de la population en Europe en fonction de leurs conditions locales (PRIMA, 2009). Un entrepôt de données permettra de comparer facilement les différents résultats (i.e. les différentes prévisions) obtenus par plusieurs versions de modèles de simulations.



a) Cube des ventes par type de produits, par magasin et par mois



b) Dimensions d'analyse

Figure 25. Exemple d'entrepôt de données (issu de (Trujillo et al., 2001))

Afin d'illustrer le principe de base des entrepôts, nous donnons ici un exemple issu de (Trujillo et al., 2001). Les **faits** d'un entrepôt sont les données à analyser (Malinowski & Zimanyi, 2008). Dans l'exemple, nous considérons que les faits de l'entrepôt de données sont les ventes de produits d'une entreprise ayant plusieurs magasins. Chaque magasin de l'entreprise fournit régulièrement ses données pour alimenter l'entrepôt. Dans un entrepôt de données, une **analyse** résulte de l'utilisation d'une opération d'agrégation (la somme, la moyenne, etc.) sur les faits. Dans l'exemple de (Trujillo et al., 2001), une analyse possible est la somme des ventes calculée par type de produits, par magasin et par mois. Le résultat de cette analyse peut être représenté dans un cube (voir la Figure 25.a). Chaque axe du cube correspond à un critère d'analyse : type de produits, magasin et mois. Les cellules du cube sont appelées des **mesures**. Elles stockent les montants des ventes pour chaque tuple <type of products, store, month>. Par exemple, dans la Figure 25.a, la somme des ventes pour le tuple <Water, Store 1, December> est égale à 4. Dans les entrepôts de données, les critères d'analyse sont structurés en hiérarchies appelées **dimension**. La Figure 25.b montre les dimensions présentées dans (Trujillo et

al., 2001). Un entrepôt de données peut produire beaucoup d'analyses en combinant différents niveaux de dimensions. Par exemple, d'autres indicateurs pourraient être calculés :

- les sommes des ventes par ville,
- les moyennes des ventes par marque, ville et année,
- les sommes des ventes par type de produits, Etat et saison,
- etc.

Les instances des niveaux (par exemple une ville, une marque, etc.) sont appelées **membres**.

Ce sont les outils OLAP qui permettront aux décideurs de choisir les dimensions à combiner et de visualiser les résultats. Les différents faits et dimensions utilisés dans un entrepôt peuvent être représentés grâce à des modèles conceptuels spécifiques. Très récemment, plusieurs travaux ont proposé des formalismes spécifiques (ER ou OO/UML) pour concevoir les modèles des entrepôts de données (appelés généralement « modèles multidimensionnels ») :

- pour les entrepôts de données non géoréférencées (Lujan-Mora, Trujillo & Song, 2006; Prat, Akoka & Comyn-Wattiau, 2006),
- pour les entrepôts de données géoréférencées (Bédard et al., 2002; Malinowski & Zimanyi, 2008).

L'objectif de ces propositions est de fournir des formalismes facilitant la spécification des faits et des dimensions.

La spécification des contraintes dans les modèles multidimensionnels reste encore très peu traitée. Il s'agit pourtant d'un problème important dans ce contexte. Différentes sources de données étant intégrées au sein des entrepôts, il est essentiel d'avoir des techniques fiables pour contrôler les contraintes d'intégrité (y compris spatiale). Les exemples de contraintes des chapitres précédents portaient sur des bases de données transactionnelles. Elles pourraient parfaitement s'appliquer sur les différentes classes décrivant un modèle multidimensionnel d'entrepôt de données. Mais par rapport à une base de données « traditionnelle » (c'est-à-dire transactionnelle), un entrepôt doit satisfaire des contraintes spécifiques relativement aux hiérarchies des dimensions. Quelques travaux décrivent quelques types de contraintes que l'on peut trouver dans les modèles multidimensionnels (Carpani & Ruggia, 2001; Ghazzi et al., 2003; Hurtado, Gutiérrez & Mendelzon, 2002; Malinowski & Zimanyi, 2008; Salehi, 2009). La récente thèse de doctorat (Salehi, 2009) présente une méthode très intéressante spécialement dédiée à la spécification des contraintes d'intégrité spatiale dans les entrepôts. Ce formalisme nous semble particulièrement adapté aux thématiciens.

Dans (Pinet & Schneider, 2009a), nous avons proposé une extension d'UML pour décrire les modèles multidimensionnels par des diagrammes de classes, ainsi qu'une représentation graphique permettant de spécifier plusieurs types de contraintes d'intégrité sur ces modèles. Par exemple, la Figure 26 présente une dimension d'analyse modélisée avec l'extension d'UML proposée. Chaque niveau de dimension est une classe ; la sémantique des flèches est celle des associations UML. La contrainte {d} indique sur le diagramme que les niveaux d'analyse « Type » et « Brand » sont disjoints c'est-à-dire que l'on va interdire de réaliser des analyses à la fois par type et par marque (par exemple calculer une somme de ventes à la fois par type et par marque). Nous avons aussi proposé une méthode qui détermine s'il est possible d'agréger correctement les faits sachant un modèle multidimensionnel et un ensemble de contraintes (voir (Pinet & Schneider, 2009a) pour plus de détails).

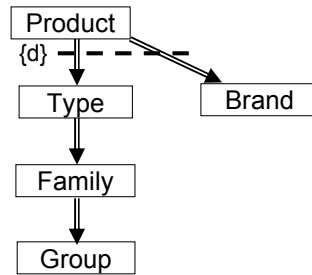


Figure 26. Une contrainte de disjonction (Pinet & Schneider, 2009a) entre deux niveaux d'une dimension

Dans le cadre des contraintes d'intégrité pour les entrepôts, nous n'avons pas trouvé dans la littérature de méthode ayant un pouvoir d'expression du niveau de celui d'un langage formel comme OCL. Nous avons donc commencé à étudier les pistes d'utilisation de notre OCL spatial pour les entrepôts de données. L'objectif du présent chapitre est de montrer qu'OCL pourrait s'adapter aux entrepôts. Il s'agit d'un chapitre plus « prospectif » associé aux travaux de la thèse de Kamal Boulil qui vient de débiter. Dans ce chapitre, nous illustrons sur une étude de cas, comment OCL (y compris spatial) pourrait être utilisé avec des modèles multidimensionnels représentés avec UML. Ces premières propositions ont été publiées dans (Pinet & Schneider, 2009a, b).

Les grandes lignes de la démarche de conception des entrepôts que nous envisageons sont présentées à la Figure 27. Cette démarche se base sur une conception avec UML et OCL, puis sur une génération automatique avec un AGL des schémas physiques des entrepôts et des mécanismes de contrôles des contraintes.

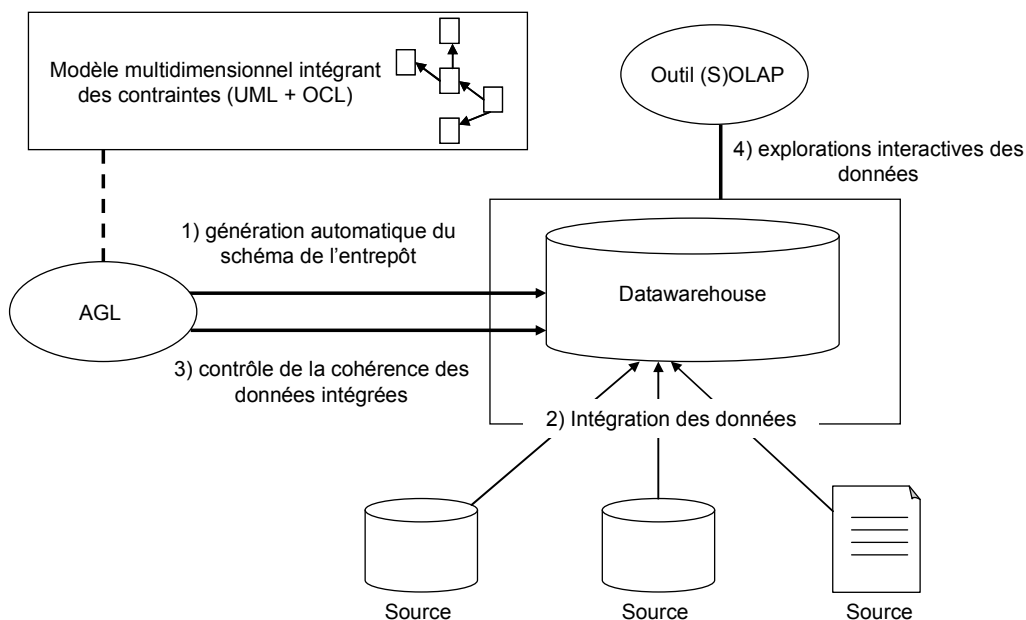


Figure 27. Etapes principales de la conception des entrepôts

La Section 5.2 explique l'exemple utilisé dans le présent chapitre. La Section 5.3 présente comment représenter en OCL, différents types de contraintes que l'on trouve couramment dans les entrepôts. La section 5.4 conclut et dresse les principales perspectives.

5.2 Présentation de l'étude de cas

Dans cette section nous introduisons une étude de cas basée sur l'exemple de la Figure 16 de la Section 3.4.1. Supposons que différentes données issues de SIGEMO soient utilisées pour produire le modèle multidimensionnel de la Figure 28. Le modèle est représenté dans le formalisme basé sur UML que nous avons introduit dans (Pinet & Schneider, 2009a).

Nous décrivons ici le modèle multidimensionnel. Dans un objectif de traçabilité, les produits à épandre (i.e. les matières organiques) sont regroupés en lots (classe *Lot de produit d'épandage*). Un lot correspond aux matières organiques provenant d'un producteur donné pour une période donnée. Les lots sont classés par catégorie de produit (classe *Catégorie de Produit*), par exemple « boue de station d'épuration ». Pour chaque lot de produit, il est précisé les différents épandages prévus (classe *Epandage*). Pour chaque épandage, il est spécifié la zone géographique sur laquelle le traitement doit être réalisé (attribut *zone d'épandage*). Une zone d'épandage est représentée par une région simple. La période durant laquelle l'épandage sera réalisé est fixée par les attributs *date de début* et *date de fin*. Les matières organiques sont classés par type (attribut *type* de la classe *Lot de produit d'épandage*) : effluents d'élevage, boues de stations d'épuration, etc. Des informations précises sur leur producteur sont aussi indiquées (classe *Producteur*).

Sur le diagramme, les flèches entre les classes correspondent à des associations plusieurs-à-un. Leur sens indique les différents niveaux des dimensions. La notation {faits} est associée à la classe des faits (i.e. les instances qui serviront aux calculs) et la notation {mesure} à l'attribut sur lequel va porter les calculs. Il y a donc 4 hiérarchies de dimension (qui partent de la classe des faits) : {Commune → Département → Région}, {Dossier}, {Lot de Produit d'Epandage → Producteur} et {Lot de Produit d'Epandage → Catégorie de Produit}. Plusieurs analyses peuvent être réalisées en combinant des niveaux de dimensions différentes.

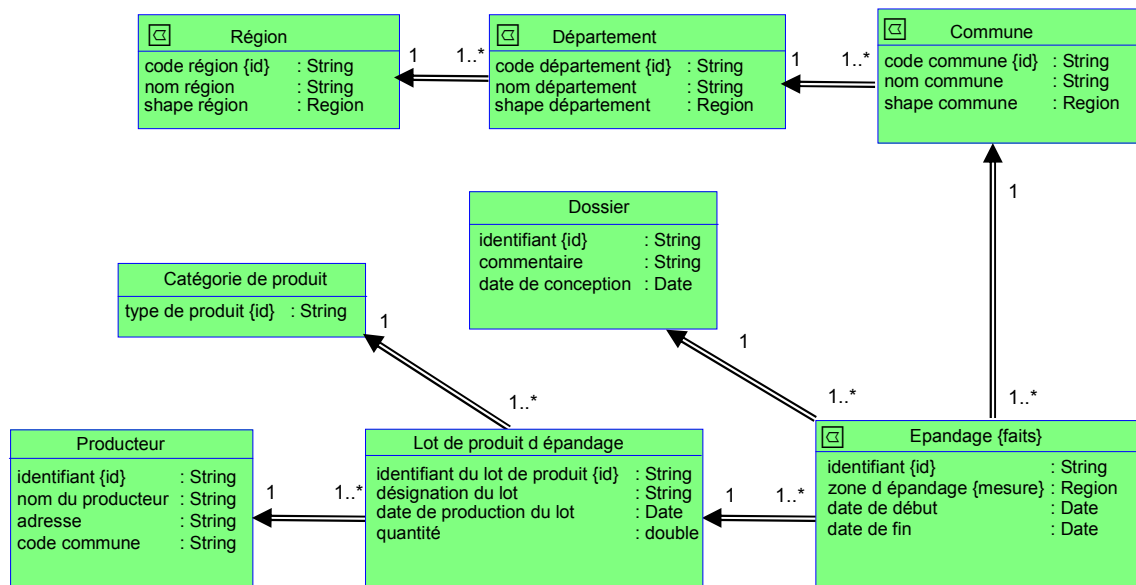


Figure 28. Modèle multidimensionnel

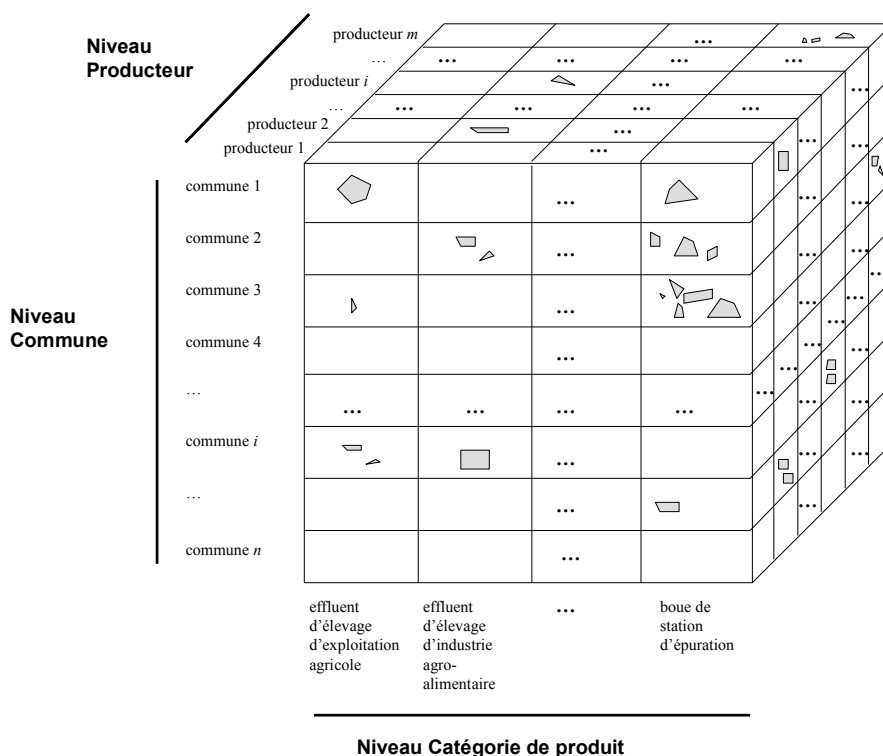


Figure 29. Une analyse présentant les résultats des zones d'épandage par producteur, commune et catégorie de produit ; la fonction d'agrégation utilisée est l'union spatiale

Considérons l'union spatiale²⁷ comme fonction d'agrégation à appliquer sur l'attribut *zone d'épandage*. Le modèle de la Figure 28 indique donc qu'il sera, entre autres, possible de calculer les zones impliquées par l'épandage (et de calculer leur surface en hectares) :

- par producteur,
- par catégorie de produit,
- par commune,
- par lots de produit d'épandage,
- par producteur et catégorie de produit,
- par producteur et commune,
- par catégorie de produit, commune et producteur,
- etc.

²⁷ L'union spatiale est une fonction classique utilisée dans les SIG - voir par exemple Manolopoulos, Y., Papadopoulos, A. & Vassilakopoulos, M. (2004) *Spatial Databases: Technologies, Techniques and Trends* IGI Global. Elles fusionnent plusieurs géométries en une seule ; la géométrie résultante est un seul objet qui peut ou non être composé de plusieurs formes disjointes.

La Figure 29 présente un cube calculable à partir du modèle de la Figure 28. Il combine les niveaux : producteur, commune et catégorie de produit.

A titre d'exemple, le Tableau 5 énumère les différentes analyses impliquant la hiérarchie spatiale {Commune → Département → Région}. L'ajout d'une nouvelle classe au modèle (e.g. lors de l'intégration d'une nouvelle source) entraînera une multiplication des analyses possibles.

| <i>Nombre de critères</i> | <i>Analyses possibles</i> |
|---------------------------|--|
| 1 | par (commune); par (département); par (région) |
| 2 | par (projet et commune); par (projet et département); par (projet et région); par (lots de produits d'épandage et commune); par (lots de produits d'épandage et département); par (lots de produits d'épandage et région); par (producteur et commune); par (producteur et département); par (producteur et région); par (catégorie de produit et commune); par (catégorie de produit et département); par (catégorie de produit et région) |
| 3 | par (catégorie de produit, producteur et commune); par (catégorie de produit, producteur et département); par (catégorie de produit, producteur et région); par (projet, lots de produits d'épandage et commune); par (projet, lots de produits d'épandage et département); par (projet, lots de produits d'épandage et région); par (projet, producteur et commune); par (projet, producteur et département); par (projet, producteur et région); par (projet, catégorie de produit et commune); par (projet, catégorie de produit et département); par (projet, catégorie de produit et région) |
| 4 | par (projet, catégorie de produit, producteur et commune); par (projet, catégorie de produit, producteur et département); par (projet, catégorie de produit, producteur et région) |

Tableau 5. Les différentes analyses spatialisées possibles à partir du modèle de la Figure 28

5.3 Exemples de contraintes

Certaines contraintes d'intégrité peuvent être contrôlées afin de vérifier s'il n'y a pas eu d'erreur dans l'intégration des données. Au niveau d'un entrepôt, une mauvaise qualité peut provenir par exemple :

- de la fusion de sources correspondant à des périodes différentes,
- d'une qualité non contrôlée dans une source de données,
- d'erreurs dans les algorithmes d'intégration,
- de la précision des données qui est différentes dans les sources,
- etc.

Quelques propositions ont été déjà faites pour formaliser les contraintes d'intégrité dans les modèles multidimensionnels (Carpani & Ruggia, 2001; Ghoszi et al., 2003; Hurtado, Gutiérrez & Mendelzon, 2002; Malinowski & Zimanyi, 2008; Salehi, 2009). Ces propositions permettent de représenter des contraintes relativement simples (par exemple indiquer des chemins alternatifs dans les hiérarchies des dimensions).

Afin de pouvoir exprimer des contraintes complexes dans les modèles multidimensionnels, nous préconisons l'utilisation d'OCL. Le fait d'employer des diagrammes de classes UML pour représenter les entrepôts (Pinet & Schneider, 2009a) nous permet d'utiliser OCL de manière assez directe. Les exemples suivants illustrent comment OCL pourrait permettre la spécification de règles d'intégrité dans les modèles multidimensionnels. Nous présentons successivement des types de contraintes importants dans les entrepôts : contraintes intra-membre, inter-membres, topologiques, temporelles, sur les hiérarchies (disjointes ou alternatives), de composition. Comme nous le montrerons à la fin de cette section, il est très facile de combiner ces différents types de contraintes de base dans une seule expression OCL afin de spécifier des conditions plus complexes. L'expression en OCL d'autres catégories de contraintes dans les entrepôts fait actuellement l'objet de la thèse de doctorat de Kamal Boulil.

Contrainte 1 (*contrainte intra-membre*)

La quantité d'un lot d'épandage est toujours strictement supérieure à zéro.

```
context Lot_de_produit_d_epandage inv :
    self.quantite > 0
```

Contrainte 2 (*contrainte inter-membres*)

Considérons maintenant une nouvelle source de données présentant la liste des exploitations agricoles. Nous pouvons intégrer cette nouvelle source avec la base de données de la Figure 28. Il est ainsi possible de rajouter un attribut booléen « est exploitant agricole » dans la classe Producteur du modèle multidimensionnel. La valeur de l'attribut est 1 si le producteur est (ou a été) exploitant agricole pendant la période considérée, et 0 sinon. Cette information est obtenue grâce à la nouvelle source présentant la liste des exploitations agricoles. Au niveau du modèle multidimensionnel, nous pouvons fixer une contrainte OCL indiquant que tout lot de produit ayant comme type « effluents d'élevage d'exploitation agricole » doit être rattaché à une exploitation. Plus formellement, si le type d'un lot est « effluents d'élevage d'exploitation agricole » alors l'attribut « est exploitant agricole » du producteur du lot doit être égal à 1.

```
context Lot_de_produit_d_epandage inv :
  if self.Categorie_de_produit.type_de_produit =
    'effluents d'élevage d'exploitation agricole'
  then self.Producteur.est_exploitant_agricole = 1
```

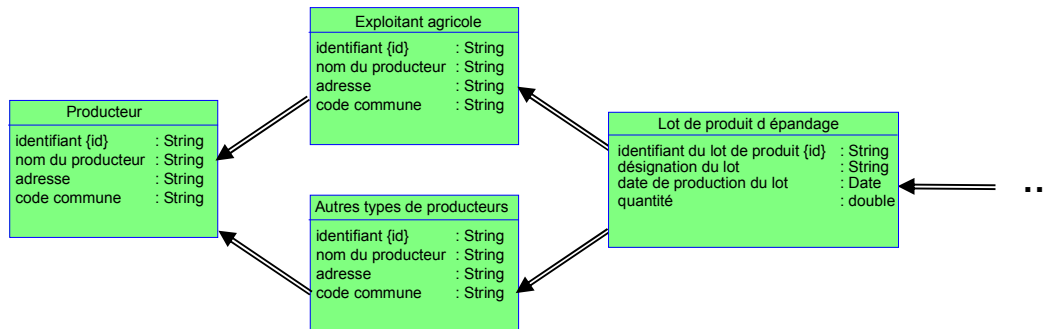


Figure 30. Deux nouvelles classes

Contrainte 3 (contrainte topologique intra-membre)

Cette contrainte indique que les communes sont soit disjointes soit adjacentes, deux à deux. Elle montre comment exploiter le langage OCL_{9IM} introduit à la Section 3.2.

```
context Commune inv :
  Commune.allInstances()->forall (C|
    (self.shape_commune) .disjoint (C.shape_commune) or
    (self.shape_commune) .meet (C.shape_commune) )
```

Contrainte 4 (contrainte topologique inter-membres)

La sémantique de cette contrainte est similaire à la contrainte 1 de la Section 3.4.2 : une zone d'épandage ne doit pas être spatialement disjointe de la commune qui lui est rattachée.

```
context Epandage inv :
  not ( (self.zone_d_epandage) .disjoint (self.Commune.shape_commune) )
```

Contrainte 5 (contrainte temporelle)

Supposons maintenant que nous souhaitions réaliser un entrepôt dédié à l'analyse des épandages d'une année donnée. Ainsi, les données qui seront chargées dans l'entrepôt ne devront concerner que l'année souhaitée. La contrainte suivante permettra de vérifier si la zone d'épandage concerne bien la période étudiée (dans l'exemple donné ici l'année 2008) ; il faut que la date de début de validité soit supérieure ou égale à 1/1/2008 et que la date de fin de validité soit inférieure ou égale au 31/12/2008. Il s'agit d'une contrainte temporelle intra-membre. Il est bien sûr possible de définir aussi des contraintes temporelles inter-membres.

```
context Epandage inv :
  self.date_de_debut >= 20080101 and
  self.date_de_fin <= 20081231
```


Contrainte 6 (contrainte de hiérarchies disjointes)

Si tous les producteurs qui sont des exploitants agricoles le restent sur toute la période considérée, alors un autre type de modèle multidimensionnel est possible. Nous pouvons introduire deux nouvelles classes : « Exploitant agricole » et « Autres types de producteurs ». Cette solution est présentée à la Figure 30. La figure ne présente que les classes qui ont été modifiées par rapport à la Figure 28. Ainsi un lot de produit est associé soit à une instance de « Exploitant agricole » soit à une instance de « Autres types de producteurs ». L'union des instances de ces deux classes correspond à l'ensemble des instances de la classe Producteur. Le modèle multidimensionnel indique donc qu'il devient possible de produire des analyses par exploitant agricole, par autres types de producteurs ou bien par tout type de producteurs. Un ensemble de contraintes peut être fixé dans ce cas. Par exemple, un lot de produit provient soit d'un exploitant agricole soit d'un autre type de producteurs. Cette contrainte peut être spécifiée par l'expression OCL suivante :

```
context Lot_de_produit_d_epandage inv :
  self.Exploitant_agricole->notEmpty() xor
  self.Autres_types_de_producteurs->notEmpty()
```

Cette dernière contrainte est importante car elle sous-entend qu'il est inutile de combiner dans une même analyse à la fois « Exploitation agricole » et « Autres types de producteurs ». Par exemple calculer les zones d'épandages par « Exploitant agricole » et « Autres types de producteur » ne donnerait aucun résultat. Cette contrainte est fréquente dans les modèles multidimensionnels (Ghozzi et al., 2003; Pinet & Schneider, 2009b). Dans (Pinet & Schneider, 2009b), elle porte le nom de contrainte « disjointe et totale ».

Contrainte 7 (contrainte de composition)

D'autres contraintes que l'on trouve fréquemment dans les modèles multidimensionnels peuvent être exprimées en OCL. Par exemple, l'ensemble des instances de la classe Producteur est composée des instances des classes « Exploitant agricole » et « Autres types de producteurs » :

```
context Producteur inv :
  Producteur.allInstances() =
  Exploitation_agricole.allInstances()->
  union(Autres_types_de_producteurs.allInstances())
```

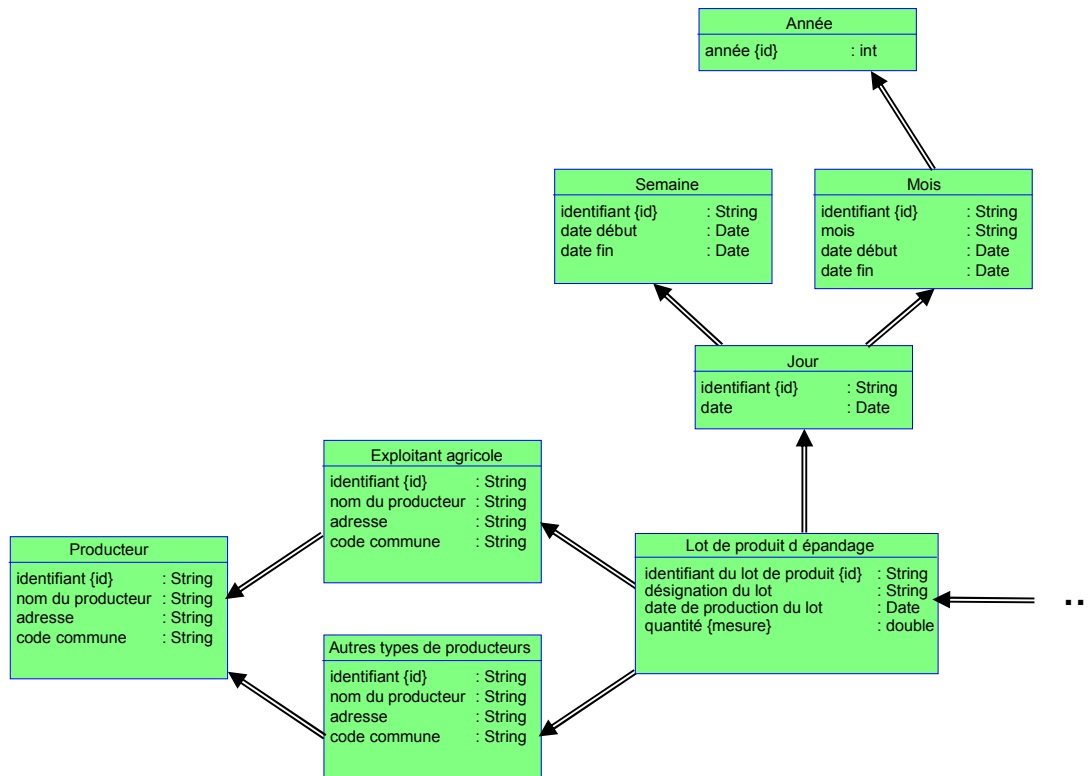


Figure 31. Ajout d'une dimension temporelle

Contrainte 8 (contrainte de hiérarchies alternatives)

Nous allons maintenant illustrer l'intégration de données calendaires dans l'entrepôt. Nous ajoutons au modèle multidimensionnel deux hiérarchies temporelles $\{\text{Jour} \rightarrow \text{Mois} \rightarrow \text{Année}\}$ et $\{\text{Jour} \rightarrow \text{Semaine}\}$. Ces hiérarchies sont reliées à la classe *Lot de produit d'épandage* grâce à l'attribut « date de production du lot » (voir Figure 31). De nouvelles analyses de données sont alors réalisables : par semaine, par jour, par mois, par semaine et par mois, par année, par semaine et par année, etc. La contrainte OCL suivante spécifie que chaque lot de produit est forcément à la fois rattaché à une semaine et à un mois.

```
context Lot_de_produit_d_epandage inv :
    self.Jour.Semaine->size() = 1 and self.Jour.Mois->size() = 1
```

OCL permet de combiner facilement différents types de contraintes de base (comme celles vues précédemment) afin d'exprimer des conditions plus complexes. Par exemple, l'expression OCL suivante est une *contrainte topologique et temporelle inter-membre*.

Contrainte 9 (contrainte topologique et temporelle inter-membres)

Dans cette contrainte, une condition temporelle implique une condition spatiale. Pour les épandages réalisées dans le 6 premiers mois de 2008, une zone d'épandage ne doit être ni disjointe, ni adjacente à la commune qui lui est rattachée.

```

context Epandage inv :
    self.date_de_debut >= 20080101 and
    self.date_de_fin   <= 20080630
    implies
    not ( (self.zone_d_epandage) .disjoint (self.Commune.shape_commune)
    or    (self.zone_d_epandage) .meet   (self.Commune.shape_commune) )
    
```

D'autres types de contraintes complexes pourraient être modélisés en OCL. Il serait par exemple possible de créer une contrainte indiquant qu'une hiérarchie n'est disjointe que pour un ensemble d'instances qui respectent une contrainte topologique donnée.

5.4 Conclusion et pistes de recherche à développer

Comme illustré dans ce chapitre, nous avons étudié les possibilités qu'offrirait OCL pour modéliser les contraintes d'intégrité dans les entrepôts de données. Cette approche vise à combler les manques dans le domaine. En effet, les problèmes liés à la modélisation et à la mise en place de procédures de contrôle de la qualité des données sont relativement peu traités dans la littérature touchant aux entrepôts. Des modélisations en OCL de contraintes typiques des entrepôts peuvent être trouvées dans (Pinet & Schneider, 2009b). Tout comme pour les bases de données traditionnelles, l'avantage d'OCL par rapport aux quelques méthodes existantes (Carpani & Ruggia, 2001; Ghazzi et al., 2003; Hurtado, Gutiérrez & Mendelzon, 2002; Malinowski & Zimanyi, 2008; Salehi, 2009) est son haut niveau d'expressivité. OCL permet de représenter des conditions complexes au sein des contraintes (voir la contrainte 9 de la section précédente). Des catégories de contraintes autres que celles présentées dans ce mémoire sont aussi spécifiques aux entrepôts. C'est par exemple le cas des restrictions des agrégations qui permettent d'indiquer qu'une certaine fonction d'agrégation est impossible pour une combinaison donnée de niveaux des dimensions (Salehi, 2009). Il faudra donc réaliser des extensions spécifiques d'OCL pour exprimer ce type de contraintes. De plus, des stratégies de génération de code adaptées doivent être trouvées.

Aussi, nous considérons dans nos travaux une spécification des contraintes au niveau du modèle multidimensionnel. Une autre approche pourrait consister à modéliser les contraintes au niveau de la description du processus d'intégration (étape 2 de la Figure 27). Des propositions ont été faites pour formaliser en UML les processus d'intégration de données (Muñoz et al., 2009; Trujillo & Luján-Mora, 2003). Ces propositions ne traitent pas de la spécification et du contrôle des contraintes au cours de ce processus.

Au-delà de la seule modélisation des contraintes d'intégrité, nous commençons à nous intéresser aussi à la modélisation conceptuelle des entrepôts spatiaux dans leur ensemble. L'objectif est de combler les besoins en termes de formalismes de modélisation. Comme l'indique (Rizzi et al.), les perspectives de recherche dans le domaine de la modélisation UML des entrepôts sont encore nombreuses et variées. A chaque niveau du développement des entrepôts et de leur utilisation, des modèles conceptuels adaptés seraient nécessaires. Quelques solutions ont été proposées par exemple pour décrire l'intégration des données dans l'entrepôt (Muñoz et al., 2009; Trujillo & Luján-Mora, 2003) ou l'exploration de l'entrepôt par des techniques de fouille de données (Zubcoff & Trujillo, 2007). Très peu de travaux portent sur les entrepôts de données spatiales. Notre objectif à moyen terme est donc de proposer une méthodologie et des formalismes efficaces pour les différentes phases de conception (intégration, conception des modèles de données, élaborations des contraintes d'intégrité, etc.). Notre approche pourra être implémentée dans OCL2SQL et dans un outil AGL, comme Objecteering, l'outil habituellement utilisé dans les projets du Cemagref. La thèse de doctorat de Kamal Boulil qui a démarré en octobre 2009 intégrera ces aspects. Les propositions de cette thèse viendront directement supporter les développements futurs concernant l'entrepôt de données du projet « SIE Pesticides ».

Toujours dans le contexte environnemental, nous étudions aussi les possibilités de concevoir un entrepôt contenant de grande quantité d'informations sur la traçabilité de la production agricole. Cette traçabilité implique un enregistrement de l'itinéraire (les trajets) des produits. Ces travaux ont été initiés dans le cadre du projet européen OTAG (Visoli et al., 2009). Le domaine d'application ciblé est la traçabilité bovine. Pour analyser les données, il est nécessaire de concevoir des « entrepôts de trajectoires ». Ce type d'entrepôt a pour objectif de stocker une grande quantité d'itinéraires des bovins en vue de les analyser. Il s'agit notamment, en cas d'épidémie touchant les bovins, de déterminer très rapidement tous les animaux présentant un risque de contamination en fonction des itinéraires qu'ils ont suivis (exploitations agricoles, pâturages, etc.). A l'heure d'aujourd'hui, il n'existe pas de travaux sur ces sujets. Nous préparons un projet de thèse sur ce thème (démarrage possible à la rentrée universitaire 2010). Le futur doctorant est Marcos Visoli actuellement ingénieur à l'Embrapa (Brésil). Marcos Visoli a commencé à travailler sur ce sujet lors de son stage de Master recherche réalisé dans notre équipe en 2009.

Une autre piste que nous explorons est la conception d'entrepôts de résultats de simulation environnementale. Concernant l'analyse de résultats de modèles de simulation environnementale, il n'existe pas d'outil dédié ayant pour but de faciliter la comparaison des résultats issus de plusieurs simulations. Cette comparaison est pourtant nécessaire afin de :

- comparer plusieurs modèles de simulation,
- comprendre l'impact de certains facteurs ; pour un ensemble de résultats de simulation de modèles choisis, calculer par exemple, à différentes échelles spatiales, la variabilité d'un facteur,
- ou bien encore dans le cas de modèles de simulation non-déterministes pour pouvoir évaluer l'impact de l'aléatoire.

Les analyses des résultats se font dans l'espace et le temps.

Aujourd'hui, l'outil SimExplorer conçu et développé au Laboratoire d'Ingénierie des Systèmes Complexes (LISC) du Cemagref, offre une interface interactive et conviviale pour exécuter les modèles numériques. Il permet de conserver une traçabilité des résultats de simulation. L'outil n'utilise pas de base de données pour conserver les résultats. Ils sont stockés dans des fichiers. La comparaison des nombreux résultats se révèle donc difficile, et nécessite l'intervention d'informaticiens. Les technologies offertes par les entrepôts de données spatiales (et temporelles) pourraient être adaptées et mises à profit pour obtenir un outil interactif d'exploration des résultats de simulation (associé à un moyen performant de stockage). Il s'agit là d'une utilisation nouvelle, jamais encore développée, et qui présente selon nous un grand potentiel. La mise en place d'une telle solution est le thème du post-doctorat de Hadj Mahboubi qui a démarré en octobre 2009 au sein du TR MOTIVE impliquant l'équipe SI et le LISC. Dans ce cadre, nous allons dans un premier temps utiliser le projet PRIMA (PRIMA, 2009) comme application.

Enfin, la piste concernant la conception d'entrepôts intégrant de grands volumes d'informations issues de capteurs présente des potentiels importants.

Conclusion et perspectives générales

Une part importante de nos travaux a porté sur la modélisation et le contrôle des règles de qualité (i.e. des contraintes d'intégrité) impliquant des relations topologiques, dans les bases de données spatiales. Cette formalisation s'intègre aux modèles de SI spatiaux (cf. Chapitre 1). Notre champ d'application concerne les bases de données agri-environnementales. La modélisation de contraintes complexes impliquant la topologie des objets était un sujet assez peu traité, mais pourtant important. En effet, de plus en plus de sources différentes sont intégrées au sein des SI agri-environnementaux. Le contrôle de la qualité des données est donc essentiel.

Ces travaux ont été supportés par deux thèses soutenues en 2007 et 2009. Dans le cadre de ces thèses, un outil opérationnel a été développé pour modéliser les contraintes d'intégrité, et générer des mécanismes de contrôle des données. Notre approche et notre outil ont été principalement utilisés dans le projet SIGEMO et STAR. Nous avons aussi expérimenté les possibilités d'appliquer notre approche sur un Système d'Aide à la Décision pour la Politique Agricole Commune (Hasenohr & Pinet, 2006; Hasenohr et al., 2005). OCL est indéniablement un bon langage pour modéliser les formes complexes des contraintes d'intégrité ciblées. Des équipes font référence à nos travaux et étendent nos propositions. Nous donnons ici quelques exemples. Faisant référence à nos travaux, l'étude publiée dans (Werder, 2009) constate la grande expressivité d'OCL pour modéliser des contraintes spatiales. Se basant sur notre approche, les travaux de (Giuseppe et al., 2009) proposent d'intégrer de nouvelles opérations spatiales à OCL. Dans (Glorio & Trujillo, 2009), les auteurs utilisent OCL_{9IM} pour modéliser des requêtes sur les entrepôts de données spatiales.

Notre activité était au départ plutôt concentrée sur les bases de données transactionnelles. Nous nous sommes plus récemment intéressés à la technologie des entrepôts de données. Nous sommes persuadés que dans un futur proche leur utilisation pourra grandement faciliter l'analyse et l'exploration des données environnementales. Nous avons donc commencé à développer des propositions de modélisation de contraintes d'intégrité (y compris spatiale) dans les entrepôts de données. Le thème de la thèse du nouveau doctorant, Kamal Boulil, intègre ce sujet de recherche.

Les trois pistes générales sur lesquels nous souhaitons avancer prioritairement sont les suivantes :

1. *Proposer des chaînes d'outils logiciels pour produire les systèmes d'aide à la décision environnementale.* S'inscrivent dans cet axe, la thèse de Kamal Boulil sur les entrepôts de données et la thèse de Guillaume Barbier sur le développement d'un outil de génie logiciel pour la simulation agronomique. Par exemple, la maîtrise de la chaîne de production des entrepôts de données spatiales permettra de concevoir plus rapidement et plus facilement des systèmes d'aide à la décision pour l'environnement. Cette chaîne pourra notamment servir au sein du Cemagref lors de projets de mise en place d'entrepôts. La Figure 32 illustre cette proposition et donne des exemples d'outils. S'il existe actuellement des outils ETL (Spatial Data Integrator, GeoKettle) et SOLAP (JMAP-SOLAP, Mondrian), il manque d'outils de type AGL facilitant la formalisation et l'implémentation des modèles multidimensionnels spatiaux. Leur mise en place nécessite les recherches que nous menons sur les extensions d'UML et OCL pour les entrepôts. En collaboration avec André Miralles (Cemagref Montpellier), la mise en œuvre d'une solution AGL adaptée pourra se faire avec Objecteering, l'AGL utilisé au Cemagref, en lien avec l'outil OCL de Dresde.

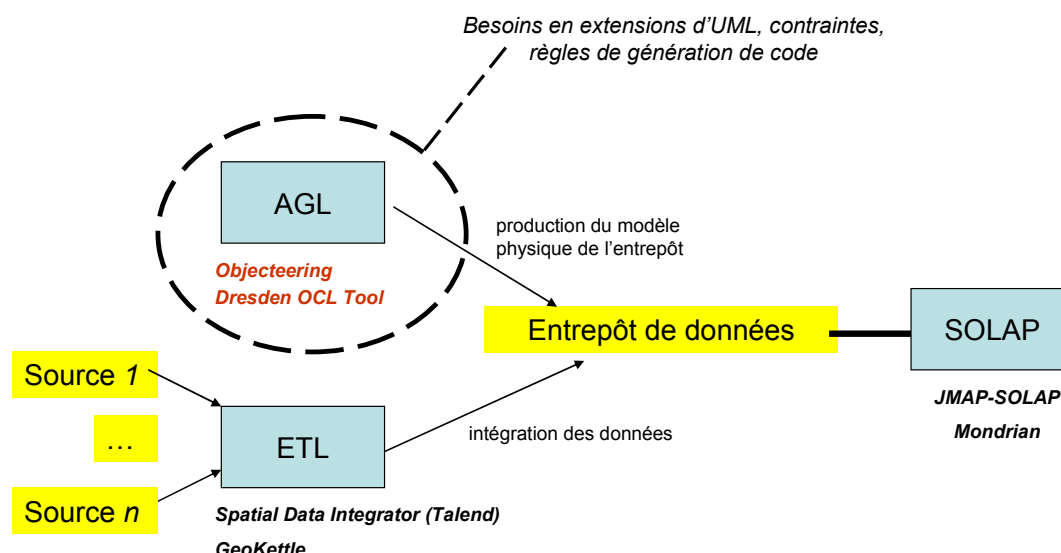


Figure 32. Exemple de chaîne de production d'entrepôt de données

- *Etudier comment les entrepôts de données peuvent être utilisés pour analyser et comparer un grand volume de résultats de simulation environnementale.* La modélisation environnementale est une activité essentielle, notamment au Cemagref où la modélisation concerne de nombreux Thèmes de Recherche. Une nouvelle piste de recherche innovante consiste à trouver les structures et les fonctionnalités que pourrait proposer ce type d'entrepôts. Le post-doctorant Hadj Mahboubi travaille actuellement sur ce sujet avec nous (Mahboubi et al., 2010).
- *Aborder de nouveaux champs d'application aux entrepôts de données dans le domaine environnemental et lever les verrous scientifiques nécessaires à leur mise en place.* Il peut s'agir d'entrepôts de données issus de réseaux de capteurs ou encore d'entrepôts de données pour la traçabilité des activités liées à l'environnement. Le projet de thèse de Marcos Visoli sur les entrepôts de trajectoires pour la traçabilité bovine (en cours de mise en place) s'inscrit dans cet axe. Cette piste de recherche s'inscrit dans la continuité de la collaboration avec les chercheurs de l'équipe SI du Cemagref travaillant sur les réseaux de capteurs (Chanet et al., 2003; Chanet et al., 2004; De Sousa et al., 2005; Kang et al., 2004; Kara et al., 2006, 2007; Visoli et al., 2008a; Visoli et al., 2008b; Visoli et al., 2009). Une thèse, encadrée dans l'équipe SI par Gil De Sousa et Jean-Pierre Chanet, portera sur la conception des SI intégrant des données de capteurs (démarrage prévu fin 2010). En collaboration avec le centre d'IBM de Montpellier et en lien avec Maguelone Tesseire (Cemagref Montpellier), nous initions aussi actuellement des travaux sur l'alimentation d'entrepôt avec des données-capteurs, ainsi que leur vérification en temps-réel par des contraintes d'intégrité.

Les différentes perspectives évoquées à la fin des chapitres 3-4-5 peuvent s'intégrer à ces trois axes : définition de nouveaux types spatiaux vagues, enrichissement du langage OCL avec de nouvelles fonctions, etc.

Le développement de ces trois axes nous semble essentiel pour les années futures. Ces axes permettront d'améliorer le traitement et l'analyse du nombre croissant de données environnementales.

Références

- Abiteboul, S., Hull, R. & Vianu, V. (2000) *Fondements des Bases de Données*.
- Abrial, J.R. (1996) *The B-Book, Assigning Programs to Meanings* Cambridge University Press.
- Akehurst, D. & Bordbar, B. (2001). On Querying UML Data Models with OCL. In «UML» 2001 — *The Unified Modeling Language. Modeling Languages, Concepts, and Tools*, pp. 91-103.
- Balley, S., Parent, C. & Spaccapietra, S. (2004) Modelling geographic data with multiple representations. *International Journal of Geographical Information Science*, **18**(4), 327 - 52.
- Bedard, Y. (2009). Web site "Spatial OLAP". In <<http://www.spatialbi.com/>>.
- Bédard, Y. (1999) Visual Modelling of Spatial Database towards Spatial PVL and UML. *Geomatica*, **53**(2), 169-85.
- Bédard, Y. (2009). Perceptory Web site. In <<http://sirs.scg.ulaval.ca/Perceptory/>>.
- Bédard, Y., Larrivee, S., Proulx, M.J. & Nadeau, M. (2004) Modeling geospatial databases with plug-ins for visual languages: A pragmatic approach and the impacts of 16 years of research and experimentations on perceptory. *Lecture Notes in Computer Science*, **3289**, 17-30.
- Bédard, Y., Pageau, J. & Caron, C. (1992). Spatial Data Modeling: The Modul-R Formalism and CASE Technology. In *ISPRS Symposium*, pp. 1-14, Washington.
- Bédard, Y. & Paquette, F. (1989). Extending entity/relationship formalism for spatial information systems. In *AUTO-CARTO 9*, pp. 818-27, Baltimore.
- Bédard, Y., Proulx, M.J., Larrivée, S. & Bernier, E. (2002). Modeling multiple representations into spatial data warehouses: A UML-based approach. In *Joint Int. Symp. ISPRS Commission IV, SDH 2002, 95th Annual CIG Conference, July, 8-12, Ottawa, Canada*.
- Behr, T. & Schneider, M. (2001) Topological relationships of complex points and complex regions. *Lecture Notes in Computer Science*, **2224**, 56-69.
- Bejaoui, L. (2009) *Qualitative Topological Relationships for Objects with Possibly Vague Shapes: Implications on the Specification of Topological Integrity Constraints in Transactional Spatial Databases and in Spatial Data Warehouses*, Université Blaise Pascal (Clermont Ferrand) / Université Laval (Québec).
- Bejaoui, L., Bédard, Y., Pinet, F. & Schneider, M. (2009) Qualitative Min-Max Model for Lines with Vague Shapes and their Topological Relations (submitted).
- Bejaoui, L., Pinet, F., Bédard, Y. & Schneider, M. (2008a) Qualified topological relations between spatial objects with possibly vague shape. *to appear in International Journal of Geographical Information Science*.
- Bejaoui, L., Pinet, F., Schneider, M. & Bedard, Y. (2008b) An adverbial approach for the formal specification of topological constraints involving regions with broad boundaries. *Lecture Notes in Computer Science*, **5231**, 383-96.
- Bejaoui, L., Pinet, F., Schneider, M. & Bedard, Y. (2010) OCL for formal modelling of topological constraints involving regions with broad boundaries. *to appear in GeoInformatica*.
- Belussi, A., Negri, M. & Pelagatti, G. (2004). GEOUML: A Geographic Conceptual Model Defined Through Specialization of ISO TC211 Standards. In *10th EC GI & GIS Workshop, ESDI State of the Art, Warsaw, Poland, 23-25 June 2004*.

Belussi, A., Negri, M. & Pelagatti, G. (2006). An ISO TC 211 Conformant Approach to Model Spatial Integrity Constraints in the Conceptual Design of Geographical Databases. In *Advances in Conceptual Modeling - Theory and Practice*, pp. 100-09.

Bernier, E., Gosselin, P., Badard, T. & Bédard, Y. (2008) Un outil Web géomatique innovateur pour mieux comprendre les risques pour la santé liés au climat. *Directions Magazine (Sep 11, 2008)* - http://francais.directionsmag.com/printer.php?article_id=2868.

Bimonte, S. (2007) *Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation*, INSA-Lyon.

Borges, K., Laender, A. & Davis, C. (1999). Spatial data integrity constraints in object oriented geographic data modeling. In *Proceedings of the 7th ACM international symposium on Advances in geographic information systems*. ACM, Kansas City, Missouri, United States.

Borges, K.A.V., Davis, C.A. & Laender, A.H.F. (2001) OMT-G: An Object-Oriented Data Model for Geographic Applications. *GeoInformatica*, **5**(3), 221-60.

Brodeur, J., Bédard, Y. & Proulx, M.J. (2000). Modelling geospatial application databases using UML-based repositories aligned with international standards in geomatics. In *Proceedings of the 8th ACM international symposium on Advances in geographic information systems*. ACM, Washington, D.C., United States.

Burrough, P.A. & Frank, A.U. (1996) *Geographic objects with indeterminate boundaries*.

Cali, A., Lembo, D., Lenzerini, M. & Rosati, R. (2003). *Source Integration for Data Warehousing. Multidimensional Databases*, pp. 361-92.

Carpani, F. & Ruggia, R. (2001). An Integrity Constraints Language for a Conceptual Multidimensional Data Model. In *SEKE'01*, pp. 220-27.

Cemagref. (2009). Site Web "Présentation du Projet X-NT". In <http://www.lyon.cemagref.fr/hh/presentations/xnt/index.shtml>.

Chanet, J.P., Boffety, D., Li, J., Pinet, F., Gerbe, P., Schneider, M., Vigier, F. & Hou, K.M. (2003). Ad Hoc network for agriculture. In *4th EFITA Conference, Debrecen-Budapest, HUN, 5-9 July 2003*.

Chanet, J.P., Jaudoin, H., Duboisset, M. & Pinet, F. (2004). Information systems technologies for sustainable management: a global approach. In *AGENG'2004 Information Systems Technologies for Sustainable Management: a Global Approach, Leuven, BEL, 11-15 September 2004*.

Claramunt, C. (2000). Extending Ladkin's algebra on non-convex intervals towards an algebra on union-of regions. In *Proceedings of the 8th ACM international symposium on Advances in geographic information systems*. ACM, Washington, D.C., United States.

Claramunt, C., Coulondre, S. & Libourel, T. (1997) Autour des méthodes orientées objet pour la conception des SIG. *Revue Internationale de Géomatique*, **7**(3-4), 237-57.

Claramunt, C. & Parent, C. (2003) Modelling concepts for the representation of evolution constraints. *Computers, Environment and Urban Systems*, **27**(3), 225-41.

Claramunt, C., Parent, C., Spaccapietra, S. & Theriault, M. (1999). Database Modelling for Environmental and Land Use. In *Geographical Information and Planning*, pp. 173-94.

Clementini, E. & Di Felice, P. (1995) A comparison of methods for representing topological relationships. *Information Sciences - Applications*, **3**(3), 149-78.

Clementini, E. & Di Felice, P. (1997) Approximate Topological Relations. *International Journal of Approximate Reasoning*, **16**(2), 173-204.

Clementini, E., Felice, P.D.I. & Califano, G. (1995) Composite regions in topological queries. *Information Systems*, **20**(7), 579-94.

Coad, P. & Yourdon, E. (1991) *Object-Oriented Analysis* Prentice Hall, Englewood Cliffs.

- Cockcroft, S. (1998). User Defined Spatial Business Rules: Storage, Management and Implementation – A Pipe Network Case Study. In *10th Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New-Zealand, 16-19 novembre 1998*, pp. 73-81.
- Cockcroft, S. (2001). Modelling Spatial Data Integrity Rules at the Metadata Level. In *6th International Conference on GeoComputation, Brisbane, Australia, 2001 september 24-26*.
- Cockcroft, S. (2004) The Design and Implementation of a Repository for the Management of Spatial Data Integrity Constraints. *GeoInformatica*, **8**(1), 49-69.
- Cohn, A.G. & Gotts, N.M. (1996). The ‘Egg-Yolk’ Representation of Regions with Indeterminate Boundaries. In *GISDATA Specialist Meeting on Spatial Objects with Undetermined Boundaries*, pp. 171-87. Taylor & Francis.
- Cook, S. & Daniels, J. (1994) *Designing object systems-object oriented modeling with Syntropy* Prentice-Hall.
- Davis, C. & Laender, A. (1999). Multiple representations in GIS: materialization through map generalization, geometric, and spatial analysis operations. In *Proceedings of the 7th ACM international symposium on Advances in geographic information systems*. ACM, Kansas City, Missouri, United States.
- De Sousa, G., Chanut, J.P., Amamra, A., Hao, J., Kang, M.A., Pinet, F. & Hou, K.M. (2005). Protocole de communication sans fil dédié : Obstacle Location-Aided Routing. In *Journée Véhicule Intelligent de l'Ecole Doctorale Sciences pour l'Ingénieur de Clermont Ferrand, 27 juin 2005*.
- Demuth, B. (2005). The Dresden OCL Toolkit and the Business Rules Approach. In *European Business Rules Conference (EBRC2005), Amsterdam*.
- Demuth, B. & Hussmann, H. (1999). Using UML/OCL Constraints for Relational Database Design. In *«UML» '99 — The Unified Modeling Language*, pp. 751-51.
- Demuth, B., Hussmann, H. & Loecher, S. (2001). OCL as a Specification Language for Business Rules in Database Applications. In *«UML» 2001 — The Unified Modeling Language. Modeling Languages, Concepts, and Tools*, pp. 104-17.
- Demuth, B., Loecher, S. & Zschaler, S. (2004). Structure of the Dresden OCL Toolkit. In *2nd International Fujaba Days “MDA with UML and Rule-based Object Manipulation”, Darmstadt, Germany, September 15 - 17*
- Devillers, R. (2004) *Conception d'un Système Multidimensionnel d'Information sur la Qualité des Données Géospatiales*, Univ. Laval – Québec / Univ. Marne-la-Vallée.
- Dilo, A. (2006) *Representation of and reasoning with vagueness in spatial information: A system for handling vague objects*. PhD Thesis, Wageningen University and ITC.
- Drummond, N. & Shearer, R. (2006). The Open World Assumption or Sometimes its nice to know what we don't Know, University of Manchester. In <http://www.cs.man.ac.uk/~drummond/presentations/OWA.pdf>.
- Duboisset, M. (2007) *Un Système de Contraintes d'Intégrité OCL pour les Bases de Données Spatiales: Application à un Système d'Information pour l'Epannage Agricole* Université Blaise Pascal.
- Duboisset, M., Pinet, F., Kang, M.A. & Schneider, M. (2005a). Integrating the calculus-based method into OCL: study of expressiveness and code generation. In *2nd IEEE International DEXA Workshop on Geographic Information Management (GIM 2005), Copenhagen, DNK, 22-26 August 2005* ;, pp. 502-06.
- Duboisset, M., Pinet, F., Kang, M.A. & Schneider, M. (2005b) Precise modeling and verification of topological integrity constraints in spatial databases: from an expressive power study to code generation principles. *Lecture Notes in Computer Science*, **3716**, 465-82.
- Duboisset, M., Pinet, F. & Schneider, M. (2007). P-OCLcbm: a language to specify topological constraints on composite regions: definition and comparison with TRCR. In *1st International*

Conference on Research Challenges in Information Science, Ouarzazate, MAR, 23-26 avril 2007, pp. 373-78.

Egenhofer, M. & Herring, J. (1992). Categorizing Binary Topological Relations between Regions, Lines, and Points in Geographic Databases. In *Technical report*. Department of Surveying Engineering, University of Maine, Orono, ME.

Egenhofer, M.J., Clementini, E. & Di Felice, P. (1994) Topological relations between regions with holes. *International Journal of Geographical Information Science*, **8**(2), 129-44.

Egenhofer, M.J. & Franzosa, R.D. (1991) Point-Set Topological Spatial Relations. *International Journal of Geographical Information Science*, **5**(2), 161-74.

Friis-Christensen, A., Tryfona, N. & Jensen, C. (2001). Requirements and research issues in geographic data modeling. In *Proceedings of the 9th ACM international symposium on Advances in geographic information systems*. ACM, Atlanta, Georgia, USA.

Gayte, O., Libourel, T., Cheylan, J.P. & Lardon, S. (1997) *Conception des Systèmes d'Information sur l'Environnement* Hermès.

Ghozzi, F., Ravat, F., Teste, O. & Zurfluh, G. (2003) Constraints and Multidimensional Databases. *ICEIS*, 104-11.

Giuseppe, P., Mauro, N., Alberto, B. & Sara, M. (2009). From the conceptual design of spatial constraints to their implementation in real systems. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. ACM, Seattle, Washington.

Glorio, O. & Trujillo, J. (2009). Designing data warehouses for geographic OLAP querying by using MDA. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 5592 LNCS, pp. 505-19.

Gubiani, D. & Montanari, A. (2008). A conceptual spatial model supporting topologically-consistent multiple representations. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*. ACM, Irvine, California.

Hadzilacos, T. & Tryfona, N. (1997) An Extended Entity-Relationship Model for Geographic Applications. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, **26**(3), 24-29.

Hasenohr, P. & Pinet, F. (2006) Modeling of a spatial DSS template in support to the Common agricultural policy. *Journal of decision systems*, **15**(2), 181-96.

Hasenohr, P., Pinet, F., Kay, S., Grenier, G. & Sevila, F. (2005). A case-study based methodology for modeling of spatial DSS using Spatial OCL. In *Joint Workshop on Decision Support Systems, Experimental Economics & e-Participation (GRAZ 2005)*, AUT, June 2005.

Hurtado, C.A., Gutiérrez, C. & Mendelzon, A.O. (2002). OLAP Dimension Constraints. In *PODS'02*, pp. 169-79.

Kang, M.A., Pinet, F., Schneider, M., Chanut, J.P. & Vigier, F. (2004). How to design geographic database? Specific UML profile and spatial OCL applied to wireless Ad Hoc networks. In *7th Conference on Geographic Information Science (AGILE'2004)*, Heraklion, GRC, April 29-May 1 2004, pp. 289-99.

Kara, M., Hou, K.M., Chanut, J.P., Zhou, H.Y., Kang, M.A. & Pinet, F. (2006). Low cost differential GPS receivers (LCD-GPS): urban vehicle tracking. In *5th IEEE International Information and Telecommunication Technologies Symposium (I2TS)*, Cuiabá, MT, Brazil, December 06-08, 2006.

Kara, M., Hou, K.M., Chanut, J.P., Zhou, H.Y., Kang, M.A. & Pinet, F. (2007). Low cost differential GPS receivers (LCD-GPS): a local cooperative differential GPS solution. In *7ème Conférence Internationale sur les NOuvelles TEchnologies de la REpartition (NOTERE'07)*, Marrakech, MAR, 4-8 juin, 2007.

- Katz, Y. & Cuenca Grau, B. (2005). Representing Qualitative Spatial Information in OWL-DL. In *Technical report*.
- Khatri, V., Ram, S. & Snodgrass, R.T. (2004) Augmenting a conceptual model with geospatiotemporal annotations. *IEEE Transactions on Knowledge and Data Engineering*, **16**(11), 1324-38.
- Khatri, V., Ram, S. & Snodgrass, R.T. (2006) On augmenting database design-support environments to capture the geo-spatio-temporal data semantics. *Information Systems*, **31**(2), 98-133.
- Klasse Objecten. (2009). OCL Tools and Services Web Site. In <<http://www.klasse.nl/ocl>>.
- Kleppe, A., Warmer, J. & Bast, W. (2003) *MDA Explained* Addison-Wesley.
- Kosters, G., Pagel, B.-U. & Six, H.-W. (1996). GeoOOA: object-oriented analysis for geographic information systems. In *Proceedings of the IEEE International Conference on Requirements Engineering*, pp. 245-53.
- Kosters, G., Pagel, B.U. & Six, H.W. (1995). Object-Oriented Requirements Engineering for GIS Applications. In *the 3rd ACM International Workshop on Advances in Geographic Information Systems, Baltimore, Maryland, December 1-2, 1995*.
- Kosters, G., Pagel, B.U. & Six, H.W. (1997) GIS-application development with GEOOOA. *International Journal of Geographical Information Science*, **11**(4), 307-35.
- Laurini, R. (2001) *Information Systems for Urban Planning: A Hypermedia Cooperative Approach* Taylor and Francis.
- Lbath, A. (1997) *AIGLE : Un environnement visuel pour la conception et la génération automatique d'applications géomatiques*, INSA de Lyon.
- Lisboa Filho, J. (2009). ArgoCASEGEO Web site. In <<http://www.dpi.ufv.br/projetos/argocasegeo/indexENG.htm>>.
- Lisboa Filho, J. & Cirano, L. (1999). Specifying analysis patterns for geographic databases on the basis of a conceptual framework. In *Proceedings of the 7th ACM international symposium on Advances in geographic information systems*. ACM, Kansas City, Missouri, United States.
- Lisboa Filho, J., Cirano, L. & Beard, K. (1998). Applying Analysis Patterns in the GIS Domain. In *the 10th colloquium of the Spatial Information Research Centre, University of Otago, New Zealand, 16-19 November, 1998*, pp. 181-88.
- Lisboa Filho, J. & Lochpe, C. (2008). Modeling with a UML Profile. In *Encyclopedia of GIS*, pp. 691-700.
- Lujan-Mora, S., Trujillo, J. & Song, I.-Y. (2006) A UML profile for multidimensional modeling in data warehouses. *Data & Knowledge Engineering*, **59**(3), 725-69.
- Mahboubi, H., Bimonte, S., Faure, T. & Pinet, F. (2010) Data warehouse and OLAP for Environmental Simulation Data. *to appear in: International Journal of Agricultural and Environmental Systems*.
- Malinowski, E. & Zimanyi, E. (2008) *Advanced Data Warehouse Design: From Conventional to Spatial and Temporal Applications*
- Mammar, A., Gervais, F. & Laleau, R. (2006). Generating B Preconditions from Typical IS Invariants. In *Technical report*. Faculté des Sciences, de la Technologie et de la Communication, Université du Luxembourg.
- Mandel, L. & Cengarle, M. (1999). On the Expressive Power of OCL. In *FM'99 — Formal Methods*, pp. 713-13.
- Manolopoulos, Y., Papadopoulos, A. & Vassilakopoulos, M. (2004) *Spatial Databases: Technologies, Techniques and Trends* IGI Global.

- Miralles, A. (2004). Méthodologie de conception à base de profils UML pour les Systèmes d'information Géographique (SIG). In *DOCTISS 2004 - La douzième session des journées des doctorants, Montpellier, 2 mars 2004*.
- Miralles, A. (2006) *Ingénierie des modèles pour les applications environnementales*. Doctorat, Informatique, Université Montpellier 2.
- Miralles, A. (2009). Ingénierie des modèles pour les applications environnementales (TETIS Web site). In http://tetis.teledetection.fr/index.php?option=com_content&task=view&id=35&Itemid=117.
- Miralles, A., Libourel, T., Papajorgji, P.J. & Pardalos, P.M. (2009). Application of a Model Transformation Paradigm in Agriculture: A simple environmental system case study. In *Advances in Modelling Agricultural Systems*, pp. 37-54. Springer.
- Miralles, A., Pinet, F. & Bedard, Y. (2010) Object-oriented modelling of spatio-temporal environmental phenomena: an overview of today's needs and solutions. *to appear in: International Journal of Agricultural and Environmental Systems*.
- Muñoz, L., Mazón, J.-N., Pardillo, J. & Trujillo, J. (2009). Modelling ETL Processes of Data Warehouses with UML Activity Diagrams. In *On the Move to Meaningful Internet Systems: OTM 2008 Workshops*, pp. 44-53.
- MurMur. (2009). MurMur Web site - Projet européen "Multi-representations and multiple resolutions in geographic databases". In <http://lbdwww.epfl.ch/e/MurMur/>.
- Nguyen, V.H., Parent, C. & Spaccapietra, S. (1997). Complex regions in topological queries. In *COSIT'97*.
- Nilakanta, S., Scheibe, K. & Rai, A. (2008) Dimensional issues in agricultural data warehouse designs. *Computers and Electronics in Agriculture*, **60**(2), 263-78.
- OMG. (2007). Unified Modelling Language: OCL, version 2.0. OMG Specification In *Technical report*.
- OpenGIS. (1999). OpenGIS Simple Features Specification for SQL, Revision 1.1. In *Technical report*.
- Pantazis, D. (1994) *Méthodologie des Phases de Conception et de Développement d'un SIG*. PhD, Université de Liège.
- Pantazis, D. & Cornélis, B. (1996) Designing and implementing a GIS in an international context. *Transactions in GIS*, **1**(4), 301-20.
- Pantazis, D. & Donnay, J.P. (1996) *La conception des SIG* Hermès.
- Papajorgji, P., Pinet, F., Miralles, A., Jallas, E. & Pardalos, P.M. (2010) Modeling: a central activity for flexible information systems development in agriculture and environment. *International Journal of Agricultural and Environmental Information Systems*, **1**(1), 1-25.
- Parent, C. (2000). A CASE Tool for Designing Spatio-Temporal Databases. In *International Workshop on Emerging Technologies for Geo-Based Applications, 22-25 May 2000, Ascona, Switzerland*, pp. 1-19.
- Parent, C., Spaccapietra, S. & Zimanyi, E. (2006) *Conceptual Modeling for Traditional and Spatio-temporal Applications* Springer.
- Parent, C., Spaccapietra, S. & Zimanyi, E. (2009). "MADS: Modeling of Application Data with Spatio-temporal features". In *MADS Web site*.
- Parent, C., Spaccapietra, S. & Zimányi, E. (2006) The MurMur project: Modeling and querying multi-representation spatio-temporal databases. *Information Systems*, **31**(8), 733-69.
- Parent, C., Spaccapietra, S., Zimanyi, E., Donini, P., Plazanet, C. & Vangenot, C. (1998). Modeling Spatial Data in the MADS Conceptual Model. In *the International Symposium on Spatial Data Handling, SDH 98, Vancouver, Canada, July 11-15, 1998*.

- Parent, C., Spaccapietra, S., Zimanyi, E., Donini, P., Plazanet, C., Vangenot, C., Rognon, N., Pouliot, J. & Crausaz, P.A. (1997) MADS: un modele conceptuel pour des applications spatio-temporelles. *Revue Internationale de Géomatique*, 7(3-4), 317-52.
- Pegoraro, C. & Velluto, M. (2007). Semantic search engine for geographic data. In *2nd Workshop COST Action C21 - Townontology*, pp. 83-99, Torino.
- Pinet, F. (2002) *Une Méthode de Conception pour la Télégéomatique : Intégration et Validation au sein d'un Atelier de Développement*, INSA Lyon.
- Pinet, F., Duboisset, M., Demuth, B., Schneider, M., Soullignac, V. & Barnabe, F. (2009a). Constraints modeling in Agricultural Databases. In *Advances in Modeling Agricultural Systems*. Springer.
- Pinet, F., Duboisset, M. & Schneider, M. (2009) Modélisation de contraintes d'intégrité spatiales avec OCL. *Revue Internationale de Géomatique*, 19(1), 93-122
- Pinet, F., Duboisset, M. & Soullignac, V. (2007) Using UML and OCL to maintain the consistency of spatial data in environmental information systems. *Environmental modelling & software*, 22(8), 1217-20.
- Pinet, F. & Lbath, A. (2003). Object-Oriented Modelling of Communicating GIS-based Applications. In *6th AGILE European Conference on Geographic Information Science, Lyon, 24-26 avril 2003*, pp. 743-50.
- Pinet, F., Roussey, C., Brun, T. & Vigier, F. (2009b). The Use of UML as a Tool for the Formalisation of Standards and the Design of Ontologies in Agriculture. In *Advances in Modeling Agricultural Systems*. Springer.
- Pinet, F. & Schneider, M. (2009a) Precise Design of Environmental Data Warehouses. *Operational Research (Springer)*, 9.
- Pinet, F. & Schneider, M. (2009b) A Unified Object Constraint Model for Designing and Implementing Multidimensional Systems. *Journal on Data Semantics*, 13, 37-71.
- Pinet, F., Ventadour, P., Brun, T., Papajorgji, P., Roussey, C. & Vigier, F. (2006). Using UML for ontology construction: a case study in agriculture. In *7th Agricultural Ontology Service (AOS) Workshop on Ontology-Based Knowledge Discovery: Using Metadata and Ontologies for Improving Access to Agricultural Information (in collaboration with AFITA 2006)*, Indian Institute of Science, Bangalore, India, 9-10 November 2006 ;, pp. 735-39. Macmillan Edition, India.
- Prat, N., Akoka, J. & Comyn-Wattiau, I. (2006) A UML-based data warehouse design method. *Decision Support Systems*, 42(3), 1449-73.
- Price, R., Tryfona, N. & Jensen, C. (1999). A Conceptual Modeling Language for Spatiotemporal Applications. In *Technical report*.
- Price, R., Tryfona, N. & Jensen, C. (2000) Extended Spatiotemporal UML: Motivations, Requirements and Constructs. *Journal of Database Management*, 11(4), 14-27.
- PRIMA. (2009). Web Site of the European Project "Prototypical Policy Impacts on Multifunctional Activities in Rural Municipalities". In <prima.cemagref.fr>.
- Proulx, M.J., Larrivée, S. & Bédard, Y. (2002). Représentation multiple et généralisation avec UML et l'outil Perceptory. In *Généralisation et Représentation multiple*, pp. 113-29. Hermes.
- Puricelli, A. (2000) *Réingénierie et Contrôle Qualité des Données en Vue d'une Migration Technologique*, INSA de Lyon.
- Reis, R., Egenhofer, M.J. & Matos, J. (2006). Topological relations using two models of uncertainty for lines. In *7th international Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, 5 - 7 July, Lisbon, Portugal, pp. 286-95.

- Rizzi, S., Abello, A., Lechtenborger, J. & Trujillo, J. Research in data warehouse modeling and design: dead or alive? . In *Proceedings of the 9th ACM international workshop on Data warehousing and OLAP*, pp. 3-10.
- Rocha, L.V.d., Edelweiss, N. & Lochpe, C. (2001). GeoFrame-T: a temporal conceptual framework for data modeling. In *Proceedings of the 9th ACM international symposium on Advances in geographic information systems*. ACM, Atlanta, Georgia, USA.
- Roussey, C., F., P., Kang, M.A. & Corcho, O. (2010). How ontologies are used for software interoperability (chapter to appear). In *Use of Ontologies to Support Information Interoperability*, Springer, 52 pages
- Rumbaugh, J., Balaha, M., Premerlani, W., Eddy, F. & Lorensen, W. (1991) *Object-Oriented Modeling and Design* Prentice-Hall.
- Salehi, M. (2009) *Developing a Model and a Language to Identify and Specify the Integrity Constraints in Spatial Datacubes (PhD Thesis)*, Université Laval.
- Salehi, M., Bédard, Y., Mostafavi, M. & Brodeur, J. (2007). On Languages for the Specification of Integrity Constraints in Spatial Conceptual Models. In *Advances in Conceptual Modeling – Foundations and Applications*, pp. 388-97.
- Saux, E., Thibaud, R., Li, K.J. & Kim, M.H. (2004). A new approach for a topographic feature-based characterization of digital elevation data. In *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems*, pp. 73-81.
- Schneider, M. (2001). A Design of Topological Predicates for Complex Crisp and Fuzzy Regions. In *Conceptual Modeling — ER 2001*, pp. 103-16.
- Schneider, M. (2008). Fuzzy Spatial Data Types for Spatial Uncertainty Management in Databases. In *Handbook of Research on Fuzzy Information Processing in Databases*, pp. 490-515.
- Schulze, C., Spilke, J. & Lehner, W. (2007) Data modeling for Precision Dairy Farming within the competitive field of operational and analytical tasks. *Computers and Electronics in Agriculture*, **59**(1), 39-55.
- Servigne, S., Ubeda, T., Puricelli, A. & Laurini, R. (2000) A methodology for spatial consistency improvement of geographic databases. *GeoInformatica*, **4**(1), 7-34.
- Shashi, S., Mark, C., Brajesh, G., Duen-Ren, L. & Shyamsundar, S. (1997) Data models in geographic information systems. *Commun. ACM*, **40**(4), 103-11.
- Shu, H., Spaccapietra, S., Parent, C. & Quesada Sedas, D. (2003). Uncertainty of Geographic Information and Its Support in MADS. In *the 2nd International Symposium on Spatial Data Quality, Hong Kong, China, March 19-20 2003*,.
- Softeam. (2009). Objecteering Web site. In <<http://www.objecteering.com/>>.
- Soullignac, V. (2000) Epannage de boues urbaines, de déchets industriels et d'effluents d'élevage sur les sols agricoles - Comparatif réglementaire en octobre 2000. *Ingénieries - Eau, Agriculture, Territoires*(24), 15-32.
- Soullignac, V., Barnabe, F., Rat, D. & David, F. (2006) SIGEMO : un système d'information pour la gestion des épandages de matières organiques. Du cahier des charges à l'outil opérationnel. *Ingénieries - Eau, Agriculture, Territoires*(47), 37-42.
- Soullignac, V., Gibold, F., Pinet, F. & Vigier, F. (2005). Spreading matter management in France within Sigemo. In *5th European Conference for Information Technologies in Agriculture (EFITA 2005)*, Vila Real, Portugal, 25-28 July 2005, p 8.
- Soullignac, V., Pinet, F. & Gibold, F. (2004). Note sur SIGEMO : Système Informatisé de Gestion des Epandages de Matières organiques. In *Congrès CASSINI 2004, Grenoble, 2-4 juin 2004*.

- Sözer, A., Yazıcı, A., Oguztüzün, H. & Tas, O. (2008) Modeling and querying fuzzy spatiotemporal databases. *Information Sciences*, **178**(19), 3665-82.
- Spaccapietra, S., Parent, C., Damiani, M.L., Macedo, J.A., Porto, F. & Vangenot, C. (2008) A conceptual view on trajectories. *Data & Knowledge Engineering*, **65**(1), 126-46.
- Spaccapietra, S., Parent, C. & Zimányi, E. (2008). Spatio-temporal and Multi-representation Modeling: A Contribution to Active Conceptual Modeling. In *Active Conceptual Modeling of Learning*, pp. 194-205.
- Tang, T. (2004) *Spatial object modeling in fuzzy topological spaces: with applications to land cover change*. PhD thesis, University of Twente.
- Trujillo, J. & Luján-Mora, S. (2003) A UML based approach for modeling ETL processes in data warehouses. *Lecture Notes in Computer Science*, **2813**, 307-20.
- Trujillo, J., Palomar, M., Gomez, J. & Song, I.Y. (2001) Designing Data Warehouses with OO Conceptual Models. *IEEE Computer*, **34**(12), 66-75.
- Tryfona, N., Anderson, S., Mogensen, S.R. & Jensen, C. (1999). A Methodology and a Tool for Spatiotemporal Database Design. In *the 7th Hellenic Conference on Informatics*.
- Tryfona, N. & Jensen, C. (2000). Using abstractions for spatio-temporal conceptual modeling. In *Proceedings of the 2000 ACM symposium on Applied computing - Volume 1*. ACM, Como, Italy.
- Tryfona, N., Pfoser, D. & Hadzilacos, T. (1997). Modeling behavior of Geographic Objects: An experience with the Object Modeling Technique. In *Advanced Information Systems Engineering*, pp. 347-59.
- Tveite, H. (2001). Structural modelling for geographical information systems using icons. In *the 8th Scandinavian Research Conference on Geographical Information Science*, pp. 185-201, Norway.
- Ubeda, T. & Egenhofer, M. (1997). Topological error correcting in GIS. In *Advances in Spatial Databases*, pp. 281-97.
- Ubeda, T. & Servigne, S. (1997). Geometrical and Topological Consistency of Spatial Data. In *Proceedings of the First International Conference on Geocomputation. 1996, Vol 1, N°1*, p. 830-842.
- Vangenot, C. (2001) *Multi-représentation dans les bases de données géographiques*, Ecole Polytechnique Fédérale de Lausanne.
- Victor de Freitas, S., Lisboa Filho, J. & Vilela, V. (2005). Improving productivity and quality of GIS databases design using an analysis pattern catalog. In *Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling - Volume 43*. Australian Computer Society, Inc., Newcastle, New South Wales, Australia.
- Victor de Freitas, S., Lisboa Filho, J., Vilela, V. & Andrade, V. (2005). Improving productivity and quality of GIS databases design using an analysis pattern catalog. In *Proceedings of the 2nd Asia-Pacific conference on Conceptual modelling - Volume 43*. Australian Computer Society, Inc., Newcastle, New South Wales, Australia.
- Visoli, M., Ternes, S., Chanet, J.P., Pinet, F., De Sousa, G. & Miralles, A. (2008a). OTAG information system. In *Technical report*.
- Visoli, M., Ternes, S., Miralles, A., Chanet, J.P., Pinet, F., De Sousa, G., Pires, P.P. & Amaral, T. (2008b). UML Model of the OTAG local system. In *Technical report*.
- Visoli, M., Ternes, S., Pinet, F., Chanet, J.P., Miralles, A., Bernard, S. & De Sousa, G. (2009). Computational architecture of OTAG project. In *EFITA 2009*, Wageningen.
- Wang, K., Fierbinteanu, C. & Maekawa, M. (2003). A Conceptual Framework for Spatiotemporal Data Modeling. In *Database and Expert Systems Applications*, pp. 57-66.
- Werder, S. (2009). Formalization of Spatial Constraints. In *12th AGILE International Conference on Geographic Information Science 2009*, Leibniz Universität Hannover, Germany.

Widom, J. & Ceri, S. (1995) *Active Database Systems: Triggers and Rules for Advanced Database Processing*

World Health Organization. (2009). WHO Global InfoBase (data warehouse). In
<<http://apps.who.int/infobase/>>.

Xinming, T. (2004) *Spatial Object Modeling in Fuzzy Topological Spaces with Applications to Land Cover Change (PhD Thesis)*, International Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands.

Yazici, A. & Akkaya, K. (2000). Conceptual Modeling of Geographic Information System applications. In *Recent Research Issues on the Management of Fuzziness in Databases*. Physica-Verlag.

Yazici, A., Zhu, Q. & Sun, N. (2001) Semantic data modeling of spatiotemporal database applications. *International Journal of Intelligent Systems*, **16**(7), 881-904.

Zadeh, L. (1965) Fuzzy sets. *Information and Control*, **8**, 338-53.

Zhong, Z., Jing, N., Chen, L. & Wu, Q. (2004) Representing topological relationships among heterogeneous geometry-collection features. *Journal of Computer Science and Technology*, **19**(3), 280-89.

Zubcoff, J.J. & Trujillo, J. (2007) A UML 2.0 profile to design Association Rule mining models in the multidimensional conceptual modeling of data warehouses. *Data & Knowledge Engineering*, **63**(1), 44-62.