

ROSEMARIE MCHUGH

INTÉGRATION DE LA STRUCTURE MATRICIELLE DANS LES CUBES SPATIAUX

Mémoire présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en sciences géomatiques
pour l'obtention du grade de maître ès sciences, (M.Sc.)

DÉPARTEMENT DES SCIENCES GÉOMATIQUES
FACULTÉ DE FORESTERIE ET GÉOMATIQUE
UNIVERSITÉ LAVAL
QUÉBEC

2008

Résumé

Dans le monde de la géomatique, la fin des années 1990 a été marquée par l'arrivée de nouvelles solutions décisionnelles, nommées SOLAP. Les outils SOLAP fournissent des moyens efficaces pour facilement explorer et analyser des données spatiales. Les capacités spatiales actuelles de ces outils permettent de représenter cartographiquement les phénomènes et de naviguer dans les différents niveaux de détails. Ces fonctionnalités permettent de mieux comprendre les phénomènes, leur distribution et/ou leurs interrelations, ce qui améliore le processus de découverte de connaissances. Toutefois, leurs capacités en termes d'analyses spatiales interactives sont actuellement limitées. Cette limite est principalement due à l'unique utilisation de la structure de données géométrique vectorielle. Dans les systèmes d'information géographique (SIG), la structure de données matricielle offre une alternative très intéressante au vectoriel pour effectuer certaines analyses spatiales. Nous pensons qu'elle pourrait offrir une alternative intéressante également pour les outils SOLAP. Toutefois, il n'existe aucune approche permettant son exploitation dans de tels outils. Ce projet de maîtrise vise ainsi à définir un cadre théorique permettant l'intégration de données matricielles dans les SOLAP. Nous définissons les concepts fondamentaux permettant l'intégration du matriciel dans les cubes de données spatiaux. Nous présentons ensuite quelques expérimentations qui ont permis de les tester et finalement nous initions le potentiel du matriciel pour l'analyse spatiale dans les outils SOLAP.

Avant-Propos

Il y a un peu plus de deux ans, je me suis engagée dans un périple qui, sans le savoir, me réservait un défi beaucoup plus grand que je ne l'aurais imaginé. Ce périple que je termine aujourd'hui n'a guère été facile et m'a fait passer par une gamme d'émotions que je ne saurais décrire. La réalisation d'une maîtrise c'est avant tout apprendre à faire face à l'inconnu et y trouver sa route. Cette étape m'a permis de me découvrir un peu plus chaque jour en tant que personne et en tant que professionnelle. J'ai appris à accepter les zones d'inconfort, à gérer mon stress, mais surtout à avoir confiance en moi et à croire en mes réalisations. La réalisation d'une maîtrise est un défi énorme que je n'aurais pas pu relever sans le support inconditionnel de nombreuses personnes. De par leurs conseils et encouragements, ces personnes m'ont tous permis de passer au travers et je souhaite sincèrement les remercier.

Tout d'abord, j'aimerais remercier mon directeur de recherche, Yvan Bédard, pour avoir cru en moi dès le début. Il m'a fait confiance et a toujours su m'encourager dans les bons moments. Je remercie aussi mon codirecteur, Frédéric Hubert, pour ses conseils et son entière disponibilité.

Je désire également remercier toute l'équipe de professionnels pour avoir répondu si patiemment à mes innombrables questions. Éveline, Sonia, Martin, Marie-Josée et Suzie, vous êtes une équipe extraordinaire!

Je ne peux passer sous silence le support de la Chaire de recherche industrielle (CRSNG) en bases de données géospatiales décisionnelles et ses partenaires pour le financement de cette recherche.

Enfin, j'aimerais remercier plus particulièrement Pascal pour m'avoir sans cesse épaulée et pour tout le soutien moral qu'il m'a apporté ainsi que mes parents et mon frère, qui ont su m'encourager tout au long de mes études.

Si je termine aujourd'hui c'est grâce à eux. Mille fois merci!

*À mes parents, Diane et Even et à
Pascal, mon partenaire de vie.
Voici le fruit de tant d'années d'étude.*

Table des matières

Résumé	i
Avant-Propos	ii
Table des matières	iv
Liste des tableaux	vii
Liste des figures.....	viii

CHAPITRE 1 Introduction..... 1

1.1 Mise en contexte.....	1
1.2 Problématique	4
1.3 Objectifs	5
1.4 Méthodologie.....	7
1.5 Synopsis.....	9

CHAPITRE 2 Revue des concepts..... 10

2.1 Introduction	10
2.2 L'information à référence spatiale.....	10
2.2.1 Types de phénomènes	10
2.2.2 Notions de couverture	11
2.2.3 Modèles conceptuels de la réalité	14
2.2.4 Structures de données géométriques	16
2.3 Gestion et exploration de l'information à référence spatiale selon une approche transactionnelle	21
2.3.1 Systèmes d'informations géographiques (SIG)	21
2.3.2 Analyse spatiale.....	22
2.3.3 Map Algebra.....	28
2.4 Gestion et exploration de l'information à référence spatiale selon une approche décisionnelle.....	31
2.4.1 Concepts OLAP	32
2.4.2 Concepts SOLAP	35
2.5 Conclusion	39

CHAPITRE 3 Concepts théoriques permettant l'Intégration du matriciel dans les cubes spatiaux.....	41
3.1 Introduction	41
3.2 Intégration du matriciel dans les dimensions spatiales.....	41
3.3 Intégration du matriciel dans les faits	47
3.3.1 Définition des types de cubes spatiaux	50
3.3.2 Choix du type de cube	51
3.4 Conclusion	53
CHAPITRE 4 Expérimentations.....	55
4.1 Introduction	55
4.2 Expérimentations exploratoires	56
4.2.1 Expérimentation exploratoire 1	56
4.2.2 Expérimentation exploratoire 2	66
4.2.3 Expérimentation exploratoire 3	74
4.3 Expérimentation formelle	77
4.3.1 Problématique actuelle.....	77
4.3.2 Approches proposées.....	78
4.4 Conclusion	85
CHAPITRE 5 Potentiel du matriciel pour l'analyse spatiale dans les outils SOLAP.....	88
5.1 Introduction	88
5.2 Possibilités d'exploitation des analyses spatiales matricielles dans un SOLAP	88
5.2.1 Les analyses obtenues par traitements au niveau des dimensions	91
5.2.2 Les analyses obtenues par traitements au niveau des mesures ..	92
5.2.3 Les analyses obtenues par traitements au niveau des mesures à la volée	93
5.2.4 Les analyses obtenues par simple navigation dans le cube	95

5.3 Conclusion99

Conclusion 101

Bibliographie 107

ANNEXE A 115

ANNEXE B 124

ANNEXE C 126

Liste des tableaux

Tableau 4-1 Relations possibles entre les cellules et les documents (Source : [Proulx et <i>al.</i> , 2008]).....	80
Tableau 5-1 Exploitation des analyses spatiales en fonction du type de cube	100

Liste des figures

Figure 1-1 Situation actuelle de l'exploitation des structures de données géométriques (vectorielle, matricielle) en fonction du type de système (SIG, SOLAP).	4
Figure 1-2 Méthodologie.....	8
Figure 2-1 Couvertures discrètes d'après ISO 19123 :2005(E).....	12
Figure 2-2 Coverage continu d'après ISO 19123 :2005(E).....	13
Figure 2-3 Modèle par objets versus modèle par couverture d'une même réalité	15
Figure 2-4 Tessellations régulières : triangulaire, carrée et hexagonale.	18
Figure 2-5 Matrice à neuf intersections d'Egenhofer 3x3 (Source : [Normand, 2002])	24
Figure 2-6 Opération de superposition spatiale en mode matriciel (Source : [Chrisman, 1997])	27
Figure 2-7 Types de fonctions de map algebra (Source : [Câmara et al., 2005])	29
Figure 2-8 Exemple d'un cube de données	33
Figure 2-9 Trois types de dimensions spatiales présentement supportés par le SOLAP (Source : [Rivest et al., 2003]).....	36
Figure 2-10 Raffinements possible du membre « Same time, Same place » (Source : [Marchand, 2004])	38
Figure 3-1 Les types de dimensions spatiales (Source : [Bédard et al., 2008])	42
Figure 3-2 Dimension géométrique matricielle	43
Figure 3-3 Dimension géométrique hybride	44
Figure 3-4 Dimension mixte matricielle.....	45
Figure 3-5 Dimension mixte hybride.....	45
Figure 3-6 Relation entre la Suisse et les cellules qui la compose (Source : [Proulx et al., 2008]).....	46
Figure 3-7 Analogie entre un modèle matriciel et un cube de données	48
Figure 3-8 Types de cubes spatiaux	50

Figure 4-1 Dimension spatiale géométrique matricielle et dimension spatiale géométrique hybride	60
Figure 4-2 Schéma du cube Muscamags	61
Figure 4-3 Relation N-N entre cellule et arrondissement.....	63
Figure 4-4 Association des points aux cellules et aux arrondissements dans la dimension géométrique hybride (découpage administratif).....	65
Figure 4-5 Nombre de ménages avec enfant(s) par arrondissement et forage au niveau cellule (500m)	66
Figure 4-6 Distance des ménages avec enfant(s) à une école	66
Figure 4-7 Dimension spatiale géométrique hybride	68
Figure 4-8 Schéma du cube TGV.....	69
Figure 4-9 Longueur et coût global des tracés potentiels de TGV.....	73
Figure 4-10 Segments du tracé Sainte-Foy 1 en zone agricole où la pente est < 4%.....	73
Figure 4-11 Schéma du cube Montmorency.....	74
Figure 4-12 Étendue cartographique des documents tels que présentés dans GÉOLAP (Source : [Proulx et <i>al.</i> , 2008]).....	78
Figure 4-13 Région mondiale représentée par les cellules de 6 degrés (Source : [Proulx et <i>al.</i> , 2008]).....	82
Figure 4-14 L'Amérique du Nord représentée par les cellules de 6 degrés, 2 degrés et 1 degré (Source : [Proulx et <i>al.</i> , 2008])	82
Figure 4-15 Filtre sur la couverture mondiale pour les documents de date de validité > 2005 (Source : [Proulx et <i>al.</i> , 2008]).....	82
Figure 4-16 Nombre de documents selon le découpage des pays (Source : [Proulx et <i>al.</i> , 2008]).....	84
Figure 4-17 Nombre de documents pour un ensemble de pays (Source : [Proulx et <i>al.</i> , 2008]).....	84
Figure 5-1 Possibilités d'exploitation des analyses spatiales vectorielles dans un SOLAP	89
Figure 5-2 Possibilités d'exploitation des analyses spatiales matricielles dans un SOLAP	90

Figure 5-3 Analyse locale de Map Algebra effectuée dans un cube matriciel et dans un modèle matriciel	94
Figure 5-4 Opération d'agrégation	95
Figure 5-5 Possibilités d'implantation du matriciel dans les dimensions spatiales permettant la visualisation de l'intersection spatiale entre deux découpages vectoriels (un niveau fin matriciel pourrait également être implanté dans la dimension géométrique vectorielle découpage santé). ..	96
Figure 5-6 Intersection spatiale entre un membre de la dimension découpage administratif et un membre de la dimension découpage santé.	97
Figure 5-7 Analyse de superposition spatiale dans un cube matriciel	98

CHAPITRE 1 INTRODUCTION

1.1 Mise en contexte

La fin des années 1990 a été marquée par l'arrivée de nouvelles solutions géomatiques décisionnelles, nommées SOLAP (*Spatial On-Line Analytical Processing*). Nés du couplage entre les technologies OLAP et les systèmes d'information géographique (SIG), les outils SOLAP fournissent des moyens efficaces pour facilement explorer et analyser des données spatiales et non spatiales à différents niveaux de détail. Un outil SOLAP est défini comme étant « un logiciel de navigation facile et rapide dans les bases de données spatiales qui offre plusieurs niveaux de granularité d'information, plusieurs thèmes, plusieurs époques et plusieurs modes de visualisation synchronisés ou non : cartes, tableaux et diagrammes » [Bédard, 2004].

Les capacités spatiales actuelles des outils SOLAP permettent de représenter cartographiquement les phénomènes et de naviguer dans les différents niveaux de détails spatiaux. L'utilisateur peut donc passer facilement d'un niveau spatial agrégé à un niveau spatial plus détaillé (ex. du niveau pays vers le niveau province), ou inversement (ex. du niveau province vers le niveau pays). En comparaison avec les technologies OLAP conventionnelles, ces fonctionnalités de représentation cartographique et de navigation dans les différents niveaux spatiaux permettent de mieux comprendre les phénomènes, leur distribution et/ou leurs interrelations, ce qui améliore le processus de découverte de connaissances. Toutefois, pour être en mesure de bénéficier pleinement de la nature spatiale des données et augmenter le pouvoir de prise de décision, il est essentiel de fournir aux utilisateurs, à même le SOLAP, certaines fonctionnalités d'analyse spatiale. Dans les systèmes d'information géographique (SIG), ces fonctionnalités permettent, entre autres, la sélection d'un sous-ensemble de la base de données en fonction de critères qui peuvent

être métriques et/ou topologiques (ex. sélectionner tous les peuplements forestiers qui sont adjacents à un cours d'eau). Du côté des outils SOLAP, la situation en termes d'analyse spatiale interactive est actuellement limitée. Cette limite est principalement due à l'usage actuel de la structure de données vectorielle dans ces outils. Il est reconnu que l'analyse spatiale vectorielle est complexe [Marchand, 2004]. En effet, en présence de grandes quantités de données ou d'analyses agrégatives typiques des applications SOLAP, elle nécessite à la fois des opérateurs complexes et lourds à exécuter ainsi que beaucoup de requêtes SQL. Tout ceci réduit donc son potentiel pour produire des réponses rapides. Or, pour conserver le fil de pensée des utilisateurs, les outils décisionnels exigent des temps de réponse très rapides lors de la navigation (< 10 secs) [Marchand, 2004; Newell, 1990]. Ainsi, étant donné la lourdeur des traitements et la complexité des algorithmes, une analyse vectorielle agrégative ou qui implique de grandes quantités de données nécessite des temps de traitements trop longs pour pouvoir être exécutée à la volée (par opposition à précalculée) dans un SOLAP. L'impossibilité d'exécuter à la volée les analyses vectorielles dans le SOLAP nécessite donc de faire un choix quant aux analyses spatiales qui doivent être précalculées afin de les rendre disponibles aux utilisateurs. Les travaux de Marchand [Marchand, 2004] proposent une méthode pour intégrer une dimension d'opérateurs topologiques (STTOD) au sein de bases de données multidimensionnelles (type de base de données utilisé pour les outils SOLAP, voir section 2.4.1). Ces opérateurs topologiques décrivent les relations spatio-temporelles possibles entre les membres des dimensions spatiales. Afin de respecter les temps de réponse exigés, cette méthode implique le précalcul et le stockage de l'ensemble des relations topologiques existantes entre les membres des dimensions spatiales. Bien que cette approche ait permis de faire un pas considérable en termes d'analyse spatiale topologique dans les outils SOLAP, la nécessité de précalculer l'ensemble des analyses limite fortement son potentiel. Ainsi, dans les travaux de Marchand [Marchand, 2004], seules les relations topologiques entre géométries linéaires simples ont été traitées, soit un faible sous-ensemble des possibilités habituelles. Malgré ceci, il a fallu 10 heures de

prétraitement à l'aide d'un Pentium 4 (768 Mb de mémoire vive) pour calculer l'ensemble des relations topologiques possibles entre 29 475 objets linéaires et 51 heures sur un ensemble de 853 764 objets. Même si cette application prenait en compte le volet temporel, il n'en demeure pas moins qu'ils étaient très loin du calcul à la volée. Ainsi, dans le but d'assurer une analyse flexible et de permettre une pleine exploration des données spatiales, l'amélioration des fonctionnalités d'analyse spatiale dans les outils SOLAP constitue une priorité qui se traduit par une augmentation du potentiel d'interactivité. Ces fonctionnalités pourraient inclure, par exemple, des analyses de corridors, de visibilité, de distances ou encore d'intersections. Cependant, la structure de données vectorielle, présentement utilisée, combinée au nombre presque infini d'analyses spatiales potentiellement d'intérêt ainsi qu'aux exigences en termes de temps de réponse limitent actuellement les possibilités d'analyses interactives dans les outils SOLAP.

Contrairement aux outils SOLAP, qui reposent seulement sur une structure vectorielle, les systèmes transactionnels exploitent deux structures de données permettant la manipulation d'informations à référence spatiale : les structures vectorielle et matricielle. En effet, la structure de données matricielle offre une alternative très intéressante au vectoriel pour effectuer certaines analyses spatiales (ex. superposition spatiale, analyse de coût, analyse de pente), ainsi que pour représenter et analyser des phénomènes continus. Cette structure de données se caractérise notamment par une topologie implicite de par la position des cellules en numéro de ligne et de colonne, ainsi que par la simplicité et la rapidité de ses opérateurs d'analyse spatiale. Le matriciel offre un potentiel supérieur en terme de rapidité, c'est pourquoi les logiciels SIG du marché ont ajouté le traitement matriciel à leurs capacités vectorielles. Malgré qu'elle possède certaines différences comparativement à une structure vectorielle notamment en termes de primitives géométriques, de résolution spatiale, de précision des résultats et d'opérateurs d'analyse, la structure matricielle offre une solution complémentaire et non compétitive à l'analyse vectorielle. Ceci est particulièrement vrai lors de l'analyse de phénomènes continus (ex. température, types de sol, altitude) ou lors du besoin de

résultats globaux, ce qui est typique de plusieurs requêtes géodécisionnelles. Conséquemment, l'analyse spatiale matricielle ne pourrait-elle pas offrir une alternative intéressante également pour les outils SOLAP? La figure 1-1 présente la situation actuelle des outils SIG et SOLAP en termes de possibilités d'exploitation des différentes structures de données géométriques.

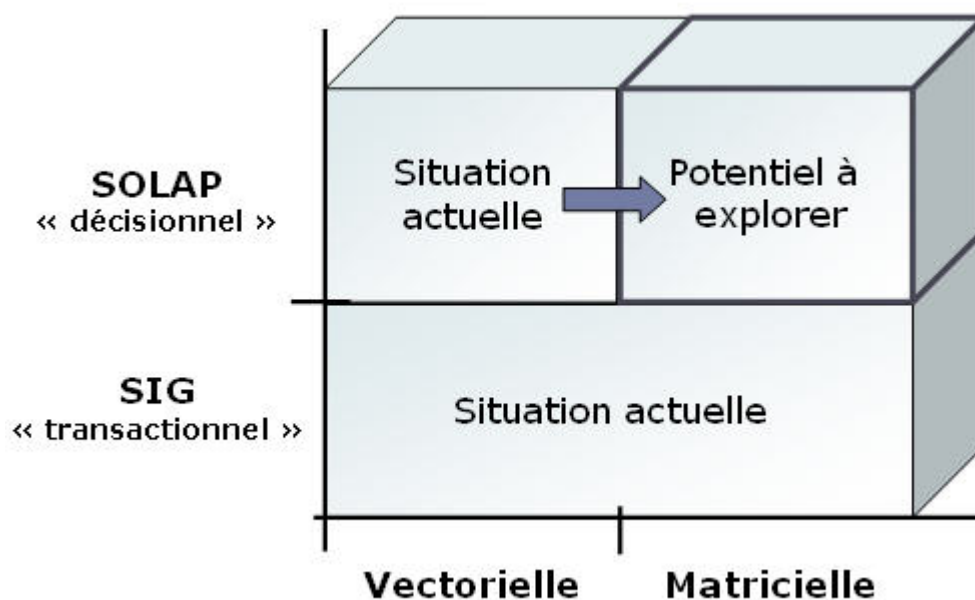


Figure 1-1 Situation actuelle de l'exploitation des structures de données géométriques (vectorielle, matricielle) en fonction du type de système (SIG, SOLAP).

1.2 Problématique

L'ajout de nouvelles fonctionnalités d'analyse aux outils SOLAP vise à apporter une plus grande possibilité de synthèse, d'agrégation et d'exploration interactive des données spatiales. L'exploitation de la structure matricielle semble offrir un potentiel très intéressant. Cependant, comme les outils SOLAP sont fondamentalement basés sur l'exploration de données par objets, l'intégration du matriciel n'est pas une procédure ou tâche évidente. Avant d'étudier le potentiel du matriciel à supporter les analyses spatiales dans un

outil SOLAP, il faut a priori déterminer comment le matriciel peut être intégré et géré dans un tel outil (étant donné que les travaux/applications existants à ce jour reposent sur la structure vectorielle). Les concepts de base sur la structure matricielle ainsi que son exploitation dans les systèmes transactionnels sont bien documentés [Bonham-Carter, 1994; Burrough et Mcdonnell, 1998; Chrisman, 1997; Collet, 1992; Tomlin, 1990; Worboys, 1995]. On retrouve également de nombreux écrits sur les concepts SOLAP et sur leur utilisation dans divers domaines tels que le transport, la foresterie et la santé [Bédard et *al.*, 1997; Bédard et *al.*, 2001; Han et *al.*, 1998; McHugh et *al.*, 2006; Rivest et *al.*, 2001]. Cependant, on ne retrouve rien dans la littérature qui évoque l'exploitation du matriciel dans des solutions du type du SOLAP, hormis les travaux de Miquel et *al.* [Miquel et *al.*, 2002]. Ces derniers utilisent le matriciel dans une application SOLAP uniquement pour répondre à une problématique bien précise d'hétérogénéités spatiale, thématique et temporelle. Leurs travaux ont exploité une structure de données matricielle sans toutefois élaborer de cadre théorique relativement aux différentes possibilités de structuration des cubes ni à son utilisation dans le SOLAP. Suite à nos réflexions sur les besoins d'améliorer le potentiel d'analyse spatiale interactive dans les outils SOLAP, nous sommes persuadés que l'exploitation d'une structure matricielle pourrait améliorer la situation actuelle de ces outils. Cependant, aucune méthode, aucune approche et aucun concept de cube matriciel n'ont été élaborés pour le moment. L'inexistence de connaissances théoriques associées à l'intégration et à la gestion de données matricielles dans les cubes de données en fait donc la problématique majeure de cette recherche qui se veut avant tout fondamentale et exploratoire.

1.3 Objectifs

L'objectif initial de cette recherche consistait à étudier le potentiel de la structure matricielle pour améliorer et accélérer les analyses spatiales dans les outils SOLAP. Toutefois, nous avons constaté en cours de route l'absence même d'une approche permettant de concevoir et d'implanter un cube avec

des données matricielles. Ce problème très fondamental ne permettait donc pas de réaliser l'étude du potentiel de l'analyse spatiale matricielle sans franchir au préalable cette étape de conception et de développement de cubes avec des données matricielles. L'évaluation des possibilités d'intégration d'une structure matricielle dans un cube de données constituait donc en soi un défi plus important, plus fondamental que ce que l'on avait imaginé. Étant basé sur une approche par objets, la possibilité d'exploiter une structure matricielle dans un cube n'était pas évidente. Elle semblait même aller parfois au-delà de ce qui est supporté par la technologie SOLAP actuelle. Ceci nous a donc amenés à nous poser quelques questions fondamentales à ce sujet :

- Doit-on exploiter la structure matricielle dans les faits ou dans les dimensions spatiales d'un cube de données?
- Peut-on exploiter les valeurs des cellules en dimension, en mesure ou en une combinaison des deux?
- Peut-on combiner des structures matricielles et vectorielles dans un même cube?
- Peut-on combiner des structures matricielles et vectorielles dans une même dimension?
- Comment peut-on bénéficier des analyses spatiales matricielles dans un cube?
- Quels impacts la structure matricielle a-t-elle sur les analyses entre dimensions spatiales?
- Et ainsi de suite...

Par conséquent, nous avons dû orienter cette recherche davantage vers les aspects théoriques de la construction de cubes matriciels, de ses possibilités, ses variantes, et son applicabilité dans un SOLAP. Notre objectif initial a donc évolué du désir d'évaluer le potentiel de l'analyse matricielle pour le SOLAP vers le désir de développer les concepts fondamentaux permettant de concevoir des cubes supportant des données matricielles, cubes qui permettront ultérieurement d'enrichir le potentiel d'analyse spatiale interactive des outils SOLAP. L'étude de ce potentiel est d'ailleurs initiée dans le chapitre 5 de ce mémoire. Notre **objectif principal** consiste donc à définir un cadre

théorique permettant l'intégration de la structure matricielle dans les cubes spatiaux.

Les objectifs secondaires de cette recherche sont :

- Décrire les possibilités d'intégrer la structure matricielle dans un cube de données
- Démontrer la faisabilité d'exploiter la structure matricielle dans un cube de données
- Initier l'étude du potentiel d'analyse spatiale matricielle en identifiant des possibilités d'exploitation de ces analyses dans les outils SOLAP.

L'établissement d'un cadre théorique pour l'intégration de la structure matricielle dans les cubes de données spatiaux permettra d'enrichir les concepts SOLAP et, ainsi, contribuera directement à l'avancement des connaissances en géomatique décisionnelle.

1.4 Méthodologie

La méthodologie suivie au cours de cette recherche comporte des étapes à la fois théoriques et expérimentales. La figure 1-2 illustre ces étapes.

Tout d'abord, une revue de littérature concernant les structures de données (vectorielles et matricielles), l'analyse spatiale, ainsi que les notions de bases de données multidimensionnelles a été réalisée. Nous avons également étudié la situation actuelle du SOLAP relativement aux possibilités d'analyse spatiale et à l'utilisation de la structure matricielle. Ensuite, nous avons effectué une synthèse de ces concepts, ce qui a notamment permis de formuler la problématique et de réviser l'objectif initialement proposé.

Ensuite, nous avons procédé à l'élaboration des concepts théoriques initiaux, principalement ceux associés aux concepts de cube matriciel et de dimensions matricielles. Cette recherche a permis d'identifier deux grandes approches complémentaires permettant d'intégrer le matriciel dans les cubes spatiaux. Les approches identifiées ont par la suite été testées et raffinées lors

d'expérimentations exploratoires. Trois expérimentations, basées sur des données réelles (ex : BNDT, forêt Montmorency) et fictives (ex : coûts fictifs associés à la construction d'un tracé de TGV), ont été réalisées à l'aide des logiciels suivants : Oracle 10g, PostgreSQL 8.2, ArcGIS 9.1, JMap et Mondrian ainsi que du langage de programmation Java.

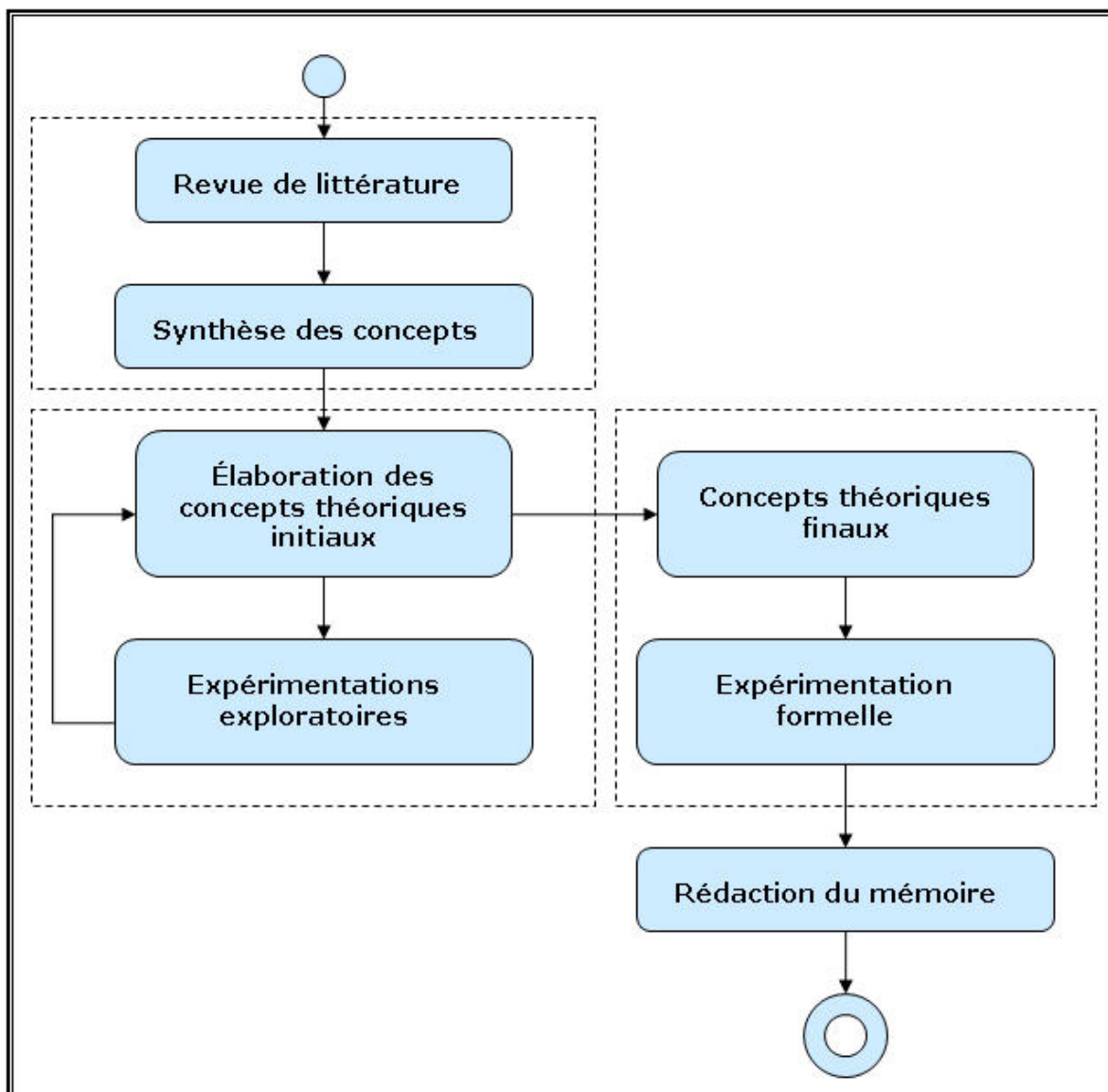


Figure 1-2 Méthodologie

Finalement, suite à la précision des concepts théoriques, nous avons testé ces derniers dans le cadre d'un projet de recherche entre la Chaire de recherche industrielle en bases de données géospatiales décisionnelles (CRSNG) et le

groupe de Recherche et Développement pour la Défense Canadienne (RDDC-Valcartier). Ce projet a notamment permis d'évaluer la pertinence et l'utilité de ces nouveaux concepts pour répondre à un besoin réel et actuel. La dernière étape de cette recherche consiste en la rédaction de ce présent mémoire. Ce qui est présenté dans ce mémoire constitue le résultat final, l'ensemble des étapes de réflexions menant à ce résultat n'y sont donc pas présentées.

1.5 Synopsis

Ce premier chapitre a exposé le contexte de cette recherche. La problématique ainsi que les objectifs ont par la suite été décrits. La méthode suivie lors de la réalisation de ce projet a ensuite été présentée. La suite de ce présent mémoire se divise en 5 chapitres.

Le chapitre suivant présente les notions théoriques qui sont à la base de ce projet de recherche. On y présente les concepts génériques de l'information à référence spatiale ainsi que son exploitation selon une approche transactionnelle et selon une approche décisionnelle. Le chapitre 3 expose les nouveaux concepts théoriques élaborés permettant l'intégration de données matricielles dans les cubes de données. Le chapitre 4 présente les différentes expérimentations exploratoires qui ont été réalisées afin de tester les concepts proposés. L'expérimentation formelle y est également décrite. Le chapitre 5 initie l'étude du potentiel matricielle pour l'analyse spatiale dans les outils SOLAP. Finalement, la conclusion de ce travail ainsi que des perspectives de recherches sont abordées dans le dernier chapitre.

CHAPITRE 2 REVUE DES CONCEPTS

2.1 Introduction

La gestion et l'exploration de l'information à référence spatiale peut s'effectuer selon deux grandes approches : l'approche **transactionnelle** et l'approche **décisionnelle**. L'approche transactionnelle est axée sur la gestion, l'intégrité et la sécurité des données, tandis que l'approche décisionnelle est orientée vers l'analyse et l'exploration. Bien qu'elles reposent sur des concepts et notions différentes, ces deux approches sont complémentaires et sont généralement utilisées conjointement dans les organisations. Ce chapitre présente, dans un premier temps, quelques concepts génériques de l'information à référence spatiale et, dans un deuxième temps, les caractéristiques spécifiques à l'approche transactionnelle et à l'approche décisionnelle, d'où sont issus respectivement les outils SIG et les outils SOLAP.

2.2 L'information à référence spatiale

Cette section présente les principaux concepts de gestion de l'information à référence spatiale utilisés dans le cadre de cette recherche. Les concepts relatifs aux types de phénomènes, à leur représentation en objets discrets ou en couverture, aux types de modélisations de la réalité ainsi qu'aux types de structures de données géométriques y sont présentés.

2.2.1 Types de phénomènes

L'information à référence spatiale permet de décrire des phénomènes du monde réel. Un phénomène est défini comme une :

Réalité qui se manifeste à la conscience, que ce soit par l'intermédiaire des sens ou non. Le phénomène constitue la réalité première. Il peut prendre la forme d'un objet (borne-fontaine), d'un événement (incendie), d'une personne (propriétaire) ou d'un concept (municipalité). [OQLF, 2004]

Il existe deux catégories de phénomènes géographiques, les **phénomènes discrets** et les **phénomènes continus**. Un phénomène discret est localisé à un endroit fixe sur le territoire et ses limites sont bien définies. Une borne-fontaine, un bâtiment et une route sont considérés comme étant des phénomènes discrets. Un phénomène continu quant à lui peut se retrouver ou être mesuré partout sur le territoire couvert [ISO, 2005; Bédard, 2003]. La température, l'altitude ou le type de sol sont des phénomènes continus. Ce sont des phénomènes géographiques qui se traduisent par une ou plusieurs variables dont les valeurs varient de façon continue dans l'espace [Donnay, 2005].

2.2.2 Notions de couverture

Compte tenu de la nature discrète de l'acquisition de l'information à référence spatiale, pour effectuer une représentation cartographique d'un phénomène continu on doit inévitablement le traduire en éléments discrets (discrétisation). La discrétisation consiste à traduire un phénomène continu par des éléments discontinus [Leterrier et *al.*, 1988]. Pour permettre la prise en compte de phénomènes géographiques continus, l'ISO [ISO, 2005] et l'OGC [OGC, 2006] ont défini une classe particulière : la couverture ou le *coverage*. La couverture est une sous-classe de la classe *feature*. Une couverture est un *feature* qui possède plusieurs valeurs par attribut, c'est-à-dire que chaque position dans la couverture possède une valeur pour chacun des attributs. Elle est caractérisée par un « domaine spatial » et une « fonction de couverture ». Le domaine spatial est représenté par une collection de géométries associées à un système de référence spatiale. C'est le domaine spatial qui définit le sous-type de la couverture. Parmi les domaines les plus souvent utilisés, on retrouve le semis de points, la grille matricielle et le TIN [ISO, 2005]. La fonction de couverture assigne pour chaque géométrie de la couverture, une valeur d'attribut ou un

vecteur de valeurs correspondant à une valeur par attribut. On distingue dans la norme deux types de couverture : (1) la **couverture discrète** et (2) la **couverture continue**.

(1) Une **couverture discrète** est formée de son domaine spatial représenté par une collection d'objets géométriques et des positions contenues dans ces objets. Sa fonction de couverture assigne pour chaque objet une seule valeur par attribut, identique pour toutes les positions d'un même objet. Un exemple de couverture discrète peut être une collection de polygones contigus pour lesquels on a assigné un type de sol unique pour chacun des polygones. La figure suivante présente les différentes sous-classes de couvertures discrètes selon l'ISO [ISO, 2005].



Figure 2-1 Couvertures discrètes d'après ISO 19123 :2005(E)

(2) Une **couverture continue** est formée d'un domaine qui consiste en un ensemble de positions spatialement localisées. Sa fonction de couverture assigne pour chaque position un vecteur de valeurs d'attributs. Dans la plupart des cas, la couverture continue est associée à une couverture discrète qui lui fournit les valeurs de contrôle permettant l'évaluation de la couverture continue. L'évaluation d'un phénomène en tout point représenté par une couverture continue est effectuée à l'aide de techniques d'interpolation. L'usage de techniques d'interpolation limite donc l'utilisation d'une couverture continue aux valeurs d'attributs quantitatives (échelles intervalle et ratio). Les valeurs qualitatives (échelles nominale et ordinale) sont de leur côté associées à une couverture discrète. La figure 2-2 présente les sous-classes de couvertures continues. Chaque sous-classe représente un modèle différent permettant le calcul des valeurs.

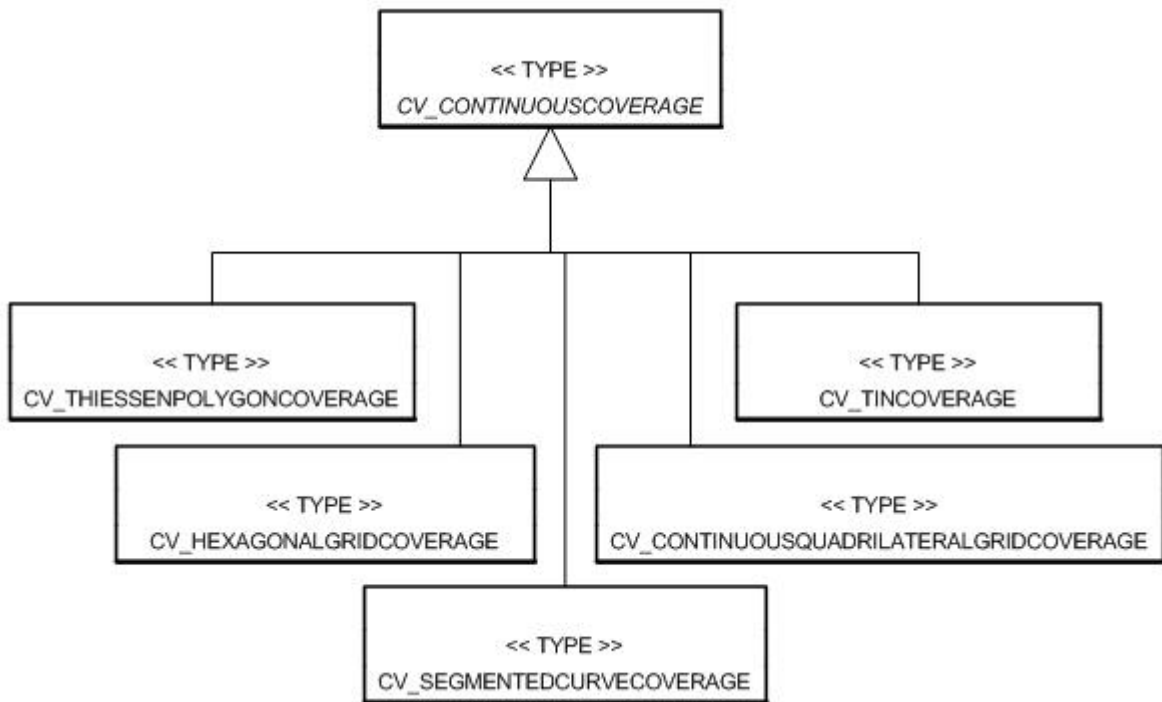


Figure 2-2 Couvertures continues d'après ISO 19123 :2005(E)

La représentation des phénomènes en objets discrets (*Feature*) ou à l'aide d'une couverture (discrète ou continue) ne sont pas mutuellement exclusives

[ISO, 2005]. C'est-à-dire que le même phénomène peut être représenté soit par un objet discret ou par une couverture. Prenons l'exemple d'une ville, cette ville peut être représentée par un objet discret polygonal pour lequel on obtient une valeur unique et uniforme pour chaque attribut (nom, superficie, population totale, densité). Cette même ville peut également être représentée par une couverture discrète composée de cellules régulières pour laquelle on obtient une valeur de population pour chaque cellule. De plus, une couverture peut être dérivée à partir d'une collection d'objets discrets qui partagent les mêmes attributs (ex. Couverture représentant la variation de la valeur marchande dérivée à partir d'une collection d'unités d'habitation possédant chacune une valeur marchande, voir figure 2-3) tout comme une collection d'objets discrets peut être dérivée d'une couverture, chaque objet étant composé d'un ensemble de positions associées à des valeurs d'attributs spécifiques (ex. agrégat de cellules matricielles qui partagent une même valeur d'attribut).

2.2.3 Modèles conceptuels de la réalité

Le choix d'une représentation en objets discrets ou en couverture (discrète ou continue) repose sur la modélisation conceptuelle qu'on se fait de la réalité. Un modèle est une représentation simplifiée d'une réalité complexe [Bédard, 2003]. L'information à référence spatiale peut être modélisée selon deux types de modèles : le *object-based model* ou le *field-based model* [Burrough et Mcdonnell, 1998; Chrisman 1975, 1978; Peuquet, 1984; Worboys, 1994, 1995]. Nous avons traduit ces concepts en **modèle par objets** et **modèle par couverture**. Ce dernier modèle décrit l'espace comme une surface continue sans limites particulières. L'information est traitée comme une collection de distributions spatiales où chaque distribution est formalisée à l'aide d'une fonction et d'un domaine spatial [Worboys, 1994, 1995]. Le modèle par objets, de son côté, décompose l'espace en une série d'objets décrits par des caractéristiques (ex : pour un bâtiment, les caractéristiques sont : le nombre d'étages, le nom du propriétaire, son type commercial, résidentiel...) et localisés à l'aide d'un système de coordonnées [Worboys, 1994, 1995]. Ces

objets ne recouvrent pas nécessairement la totalité de l'espace et peuvent se superposer. Considérons, par exemple, l'observation d'unités d'habitation, suivant leur localisation et leur valeur marchande en tant que caractéristiques. Dans un modèle par objets, on distingue des objets (unités d'habitation) localisés avec, pour chacun, leur valeur marchande. Dans un modèle par couverture, on représente de manière continue la valeur marchande à l'aide d'une couverture discrète ou continue. La figure suivante schématise les deux types de modélisations pour l'exemple ci-haut.

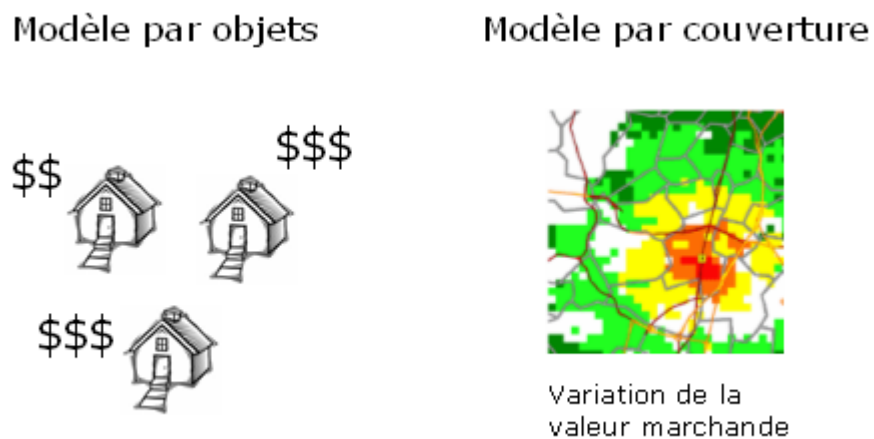


Figure 2-3 Modèle par objets versus modèle par couverture d'une même réalité

Des structures de données sont ensuite utilisées pour formaliser ces modèles conceptuels de manière à pouvoir enregistrer et manipuler l'information à référence spatiale à l'aide de systèmes informatiques. Les deux principales approches utilisent une série de points, lignes et polygones, ou une tessellation pour décrire les différents phénomènes de l'espace. L'utilisation d'un modèle par objets ou par couverture aura donc une influence sur la structure de données qui sera utilisée pour décrire les phénomènes et sur les analyses qui pourront être réalisées [Burrough et McDonnell, 1998].

2.2.4 Structures de données géométriques

La formalisation de l'information à référence spatiale s'effectue principalement à l'aide de deux structures fondamentales : (a) une structure vectorielle et (b) une structure en tessellation. La structure employée influence la manière dont l'information est décrite, structurée et stockée dans un système informatique. Nous présentons ici ces deux grandes catégories de structures de données. La structure matricielle dérivée de celle en tessellation est également abordée en (c) et nous terminons cette section par un résumé des principaux avantages et inconvénients des structures vectorielles et matricielles.

(a) Structures vectorielles

Les primitives géométriques des structures vectorielles sont le point, la ligne et le polygone (et le volume dans le cas de systèmes en 3D). Ces primitives sont stockées dans une ou plusieurs tables d'une base de données, lesquelles contiennent des séries de coordonnées et dans certains cas, les relations qui existent entre les primitives. Ces structures permettent de représenter des occurrences d'objets (objets discrets). Chaque occurrence géographique représente un objet de la réalité possédant un ensemble d'attributs qui lui est propre. Les relations spatiales entre les objets sont déduites mathématiquement ou à l'aide de la topologie. Les structures vectorielles peuvent être classifiées par un spectre, allant du modèle spaghetti vers un modèle topologique ce qui équivaut à aller d'une structure efficace et rapide pour le dessin vers une structure flexible pour la gestion des objets spatiaux [Bédard, 2003].

Dans la **structure spaghetti**, les éléments géométriques sont définis par leurs coordonnées (ex : x , y et z). Un point est défini par une paire de coordonnées (x , y), une ligne est définie par au moins deux paires de coordonnées à ses extrémités et un polygone est défini par les paires de coordonnées de ses sommets. Dans cette structure, seules les classes géométriques sont stockées et les coordonnées qui constituent l'objet. Les relations spatiales entre les classes d'objets ne sont pas codées. Les relations topologiques sont donc

déduites des coordonnées et il existe conséquemment beaucoup de redondance des données.

Dans une **structure de type topologique**, les entités géométriques de bases sont le nœud, la chaîne et le polygone. Une chaîne est définie par un nœud à chaque extrémité, un nœud étant la jonction entre au moins 2 chaînes. Le polygone est défini, quant à lui, par des chaînes qui forment son enveloppe. Dans cette structure, une chaîne commune à deux polygones adjacents n'est conservée qu'à un seul endroit dans la base de données, indépendamment de l'objet ou des objets auxquels elle appartient. Contrairement au modèle spaghetti où cette chaîne est conservée en double. Cette structure conserve de façon explicite les relations entre les trois entités de base. C'est la méthode la plus utilisée pour coder les relations spatiales. La structure topologique permet de faire certaines analyses spatiales simplement en naviguant dans la base de données, sans utiliser les coordonnées, ce qui évite de longs calculs pour effectuer ces analyses. Par contre, sa structure est plus complexe et demande un temps considérable de construction.

(b) Structures en tessellation

Une structure en tessellation correspond à un découpage continu de l'espace en une collection de polygones contigus. Ces polygones peuvent être de forme régulière ou irrégulière.

Une **tessellation irrégulière** est un découpage de l'espace à partir d'un ensemble de polygones différents [Bédard, 2003]. Les tessellations irrégulières les plus utilisées par les SIG sont : le TIN (*Triangulated Irregular Network*), la triangulation de Delaunay et le diagramme de Voronoi. D'autres tessellations comme les tessellations en rectangles, ou encore les tessellations qui suivent les méridiens et les parallèles sur le globe (feuillet cartographique) font également parti de cette catégorie de tessellation.

Une **tessellation régulière** est un découpage de l'espace formé de polygones réguliers de même forme et de même taille [Worboys, 1995]. Un polygone est

considéré comme régulier si tous ses côtés sont de même longueur et si tous ses angles intérieurs sont égaux. Dans un espace cartésien, il y a seulement trois tessellations régulières possibles : en triangles, en carrés ou en hexagones. Ces trois tessellations sont présentées à la figure 2-4. La tessellation régulière en carrés est de loin la plus utilisée. Elle est plus communément appelée structure matricielle ou structure raster. La tessellation en carrés est la seule qui peut être décomposée ou agrégée en objets de même forme et de même orientation [Laurini et Thompson, 1992]. Les cellules carrées offrent une plus grande flexibilité dans le sens où l'ensemble des algorithmes d'analyse spatiale en mode matriciel sont basés sur ce type de tessellation [Tomlin, 1990]. La structure matricielle fait l'objet d'une description plus approfondie dans le paragraphe suivant.

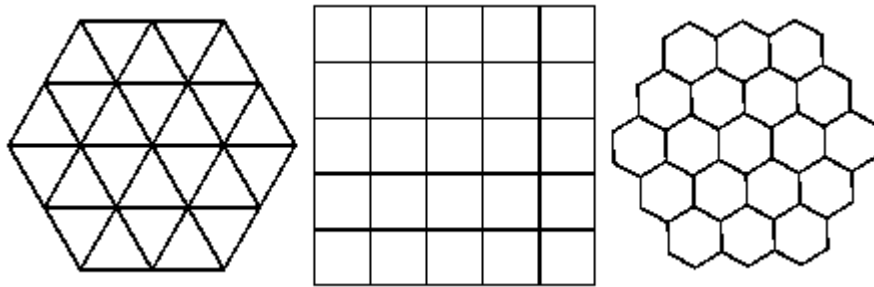


Figure 2-4 Tessellations régulières : triangulaire, carrée et hexagonale.

(c) Structure matricielle

La structure matricielle est la plus simple d'utilisation. Elle divise l'espace en une grille de polygones réguliers de forme carrée, communément appelée cellule, à chacune desquelles sont associés un ou plusieurs attributs. Ces cellules représentent la primitive géométrique de la structure matricielle. La grille qui forme un modèle matriciel représente une subdivision continue et régulière de l'espace. La position de chaque cellule est identifiée par un numéro de ligne et de colonne dans la grille. Les coordonnées des cellules ne sont pas explicitement stockées puisque l'identification en numéro de ligne et de colonne le fait implicitement. La taille des cellules définit la résolution

spatiale de l'information géographique. Le choix de la résolution est fonction de la précision avec laquelle on veut représenter les entités. La résolution a un impact direct sur le volume de données généré. Ce choix doit donc être un compromis entre représenter le plus fidèlement la réalité et minimiser le volume de données. Puisque l'espace mémoire requis augmente considérablement lorsque la résolution devient de plus en plus fine, il est possible d'utiliser des méthodes de compression comme le *quadtree* (lorsqu'utilisé comme méthode de compression) ou le *run-length* pour pallier au problème d'espace mémoire [Burrough et McDonnell., 1998; Worboys, 1995]. Dans une structure matricielle, la composante thématique est véhiculée par les valeurs numériques de la grille, tandis que la composante spatiale est implicitement décrite par la position relative en ligne et colonne de chaque cellule. L'ensemble des cellules formant la zone de saisie et leur valeur d'attribut associée est appelé une couche. Un modèle matriciel peut contenir plusieurs couches où chacune d'elle correspond à un attribut particulier de la zone d'étude (le type de sol, la température, l'altitude, etc.) ou peut contenir une seule couche contenant plusieurs attributs. Dans un tel modèle, la superposition de couches est une opération très simple et très rapide qui permet entre autres d'effectuer une sélection de cellules répondant à des critères particuliers. La structure matricielle permet de représenter dans une même grille tout type d'objets géométriques (point, ligne, polygone), ce qui implique que des données spatiales de différents types peuvent facilement être superposées sans nécessiter des calculs géométriques complexes [Bonham-Carter, 1994]. Un point est représenté par une seule cellule, une ligne par une série de cellules alignées, et un polygone par un ensemble de cellules contiguës. Bien qu'il soit possible de représenter des objets géographiques discrets à l'aide d'une structure matricielle, cette structure est mieux adaptée à la représentation de phénomènes continus [Worboys, 1995].

Structure vectorielle ou structure matricielle?

Ces deux structures (vectorielle, matricielle) possèdent chacune leurs avantages et inconvénients. L'intérêt de la structure vectorielle réside dans son

efficacité à identifier chaque objet géographique individuellement et à décrire seulement l'espace qu'il occupe, ce qui résulte en une représentation discrète et fine de la réalité. Par contre, les opérateurs d'analyse spatiale en mode vectoriel vont souvent être complexes et les temps de traitements peu rapides exceptés les analyses sur des objets (ex. Count). La structure matricielle est, de son côté, idéale pour représenter des phénomènes continus. Sa structure est simple car seul le territoire est discrétisé de façon régulière, indépendamment des objets géographiques, et les relations spatiales y sont implicites grâce à la position des cellules exprimée en numéro de ligne et de colonne dans une grille. Inversement à la structure vectorielle, ses opérateurs d'analyse spatiale sont simples, car basés sur un territoire régulier et les temps de traitements sont rapides alors que les opérations sur des objets (ex. Count) sont complexes et longues. La structure matricielle présente l'inconvénient d'occuper beaucoup d'espace mémoire et elle représente la réalité de façon plus grossière (dépendamment de la résolution, laquelle n'atteint toutefois jamais celle de la représentation vectorielle pour des raisons d'efficacité). La question de la résolution est toujours un compromis dans le cas d'une structure matricielle. En effet, si la résolution est trop grande, le volume de donnée est également trop grand et non significatif. C'est-à-dire que la variation des valeurs d'attributs est peu fréquente d'une cellule à l'autre, ce qui n'est pas le cas des résolutions grossières. Conséquemment, dans les cas où la résolution est trop fine, il vaut mieux utiliser une structure vectorielle pour laquelle seuls les éléments significatifs sont stockés.

En dépend des avantages et inconvénients de chacune des structures, le choix de la structure de données dépend en grande partie de la modélisation conceptuelle qu'on se fait de la réalité. Une modélisation par objets résulte typiquement en une utilisation d'une structure vectorielle et les phénomènes discrets et continus sont représentés par des objets discrets. Une modélisation par couverture résulte habituellement en une utilisation de la structure matricielle et la représentation des phénomènes se fait à l'aide d'une couverture [Worboys, 1994, 1995] (cf. section 2.2). Les concepts de modèles conceptuels et de structures de données sont donc souvent liés. Le choix de la

structure de données peut également être en fonction des entités à gérer, de la résolution spatiale de représentation géométrique, du type des traitements à effectuer, du niveau de performance requis et des besoins en termes de gestion de l'intégrité des données. Le choix a également une grande influence sur le type d'analyse qu'il est possible de réaliser. Selon Tomlin [Tomlin, 1990], le matriciel serait plus adapté à l'interprétation du **où** et le vectoriel à l'interprétation du **pourquoi**.

2.3 Gestion et exploration de l'information à référence spatiale selon une approche transactionnelle

2.3.1 Systèmes d'informations géographiques (SIG)

À l'heure actuelle, les données spatiales sont, dans la grande majorité des cas, gérées et exploitées par des outils de type transactionnel. Ces outils sont optimisés pour la saisie, le stockage, la mise à jour, la sécurité et l'intégrité des données et ont été conçus pour gérer des transactions courantes. Les systèmes d'informations géographiques (SIG) sont des exemples d'outils transactionnels. Selon le domaine d'application, il existe dans la littérature plusieurs définitions du terme SIG. Worboys [Worboys, 1995] et Bédard [Bédard, 2003] le définissent de manière générale comme un système informatique capable de stocker, assembler, manipuler et afficher l'information à référence spatiale. Dans les SIG, l'information à référence spatiale peut se décomposer en trois grandes composantes, une composante thématique, une composante spatiale et une composante temporelle [Worboys, 1995]. La composante thématique porte sur les propriétés sémantiques du phénomène (ex. nom d'une ville, sa population, etc.). La composante spatiale est elle-même décomposée en deux sous-composantes : une géométrique et une topologique. La composante géométrique renseigne sur la position ou la forme [OQLF, 2004] et la composante topologique sur les relations avec les autres phénomènes (adjacence, inclusion, connectivité...). Finalement, la composante

temporelle permet de suivre l'évolution de l'information thématique et spatiale en fonction du temps.

Les fonctionnalités des SIG permettent de manipuler à la fois des données structurées en vectorielle et en matricielle. La conversion d'une structure à l'autre est également réalisable dans ces outils. L'opération de rasterisation permet de convertir un thème vectoriel en format matriciel. La vectorisation permet l'inverse, soit de convertir un thème matriciel en format vectoriel.

Les SIG offrent de nombreuses fonctionnalités d'analyse spatiale permettant ainsi une pleine exploration de l'information à référence spatiale et la découverte de nouvelles connaissances. La section suivante présente le concept d'analyse spatiale et fait le point sur les types d'analyses qui sont mieux adaptés pour une structure de données matricielle versus ceux mieux adaptés pour une structure de données vectorielle.

2.3.2 Analyse spatiale

L'analyse spatiale est un « raisonnement qui permet de déduire les caractéristiques d'un phénomène en faisant intervenir des données à référence spatiale » [OQLF, 2004]. L'analyse spatiale regroupe un ensemble de méthodes et d'outils qui permettent de comprendre, évaluer et interpréter la répartition spatiale des phénomènes dans le but de découvrir et/ou de mettre en évidence des règles générales d'organisation de l'espace [Pumain et Saint-Julien, 2004]. Une analyse spatiale peut être appliquée à l'interrogation des composantes thématiques, géométriques ou topologiques de l'information à référence spatiale contenue dans les SIG. Il existe un grand nombre d'opérateurs élémentaires d'analyses spatiales, mais, à notre connaissance, il n'existe pas de classification formelle de ces opérateurs. Afin de présenter les principaux opérateurs utilisés dans les SIG, nous proposons une classification de ceux-ci. Nous distinguons premièrement deux grandes catégories d'analyses, les analyses simples et les analyses avancées. Les analyses simples sont réalisées à l'aide d'opérateurs (1) thématiques, (2) métriques ou (3)

topologiques. Tandis que les analyses avancées sont réalisées à l'aide d'une combinaison d'opérateurs simples. Parmi les analyses avancées nous distinguons les opérateurs (4) d'analyse 3D, (5) d'analyse de proximité et finalement de (6) superposition spatiale (*Spatial Overlay*). Cette classification n'est pas une fin en soi pour le présent mémoire et ne sera pas davantage élaborée, elle permet simplement de présenter les principaux opérateurs d'analyse disponibles dans les SIG.

(1) Opérateurs d'analyse thématique

Les opérateurs d'analyse thématique sont des opérateurs qui agissent sur la composante thématique de l'information à référence spatiale. C'est-à-dire que les opérations sur les objets spatiaux sont indépendantes des composantes géométriques et topologiques de ces derniers. En mode vectoriel, ces analyses sont réalisées à l'aide du langage SQL et permettent de sélectionner des objets selon leurs attributs (ex. Sélection des unités d'habitation dont la valeur marchande est supérieure à 150 000 \$) [Proulx, 1995]. En mode matriciel, Collet [Collet, 1992] définit les outils d'analyse thématique comme des opérateurs qui transforment chaque cellule indépendamment de ses voisins et de sa position dans la grille. Ces transformations peuvent être univariées ou multivariées, c'est-à-dire qu'elles impliquent une seule couche ou plusieurs couches à la fois. Elles font appel à des opérateurs logiques, arithmétiques, mathématiques ou encore statistiques et sont appliquées individuellement à chacune des cellules de façon successive. Les fonctions locales sont considérées comme étant des opérateurs d'analyse thématique (voir la section 2.3.3 sur le Map algebra pour plus de détails).

(2) Opérateurs d'analyse métrique

Par définition, les analyses métriques portent sur ce qui se « mesure » (c.-à-d. nécessite une unité de mesure), soit sur la composante géométrique (position, forme, orientation, taille) des objets spatiaux. Calculer la superficie d'un terrain, mesurer la longueur d'une route et calculer le centroïde d'un polygone sont des opérations d'analyse métrique. Ces analyses sont généralement

effectuées sur des objets discrets, la structure vectorielle est donc plus appropriée pour ce type d'analyse si l'on cherche des valeurs précises. Il est possible de réaliser des analyses métriques en matriciel, toutefois si on veut appliquer ces fonctions à des objets discrets, les objets spatiaux doivent préalablement être identifiés. C'est-à-dire que pour mesurer la longueur d'une route, on doit identifier, au préalable, les cellules qui composent cette route. Par contre, la structure matricielle est avantageuse pour faire des calculs métriques sur l'ensemble du territoire couvert, par exemple pour obtenir la superficie totale couverte par une forêt ou la proportion couverte par des plans d'eau.

(3) Opérateurs d'analyse topologique

Par définition, les analyses topologiques portent sur la composante invariante des relations entre objets spatiaux qui demeure vraie malgré les transformations géométriques du système de référence spatiale où sont localisés les objets. Les relations topologiques entre un objet A et un objet B sont définies par l'intersection de l'intérieur, de la limite et de l'extérieur de ces objets. Le modèle à neuf intersections d'Egenhofer [Normand, 2002], illustré par la figure 2-5, a été choisi pour représenter toutes les relations topologiques possibles entre objets. Ce modèle est reconnu par les principaux organismes internationaux de normalisation de l'information géographique : l'Open Geospatial Consortium (OGC) (www.opengeospatial.org) et l'ISO TC 211 (www.isotc211.org).

	Intérieur de B	Limite de B	Extérieur de B
Intérieur de A	$\emptyset/\neg\emptyset$	$\emptyset/\neg\emptyset$	$\emptyset/\neg\emptyset$
Limite de A	$\emptyset/\neg\emptyset$	$\emptyset/\neg\emptyset$	$\emptyset/\neg\emptyset$
Extérieur de A	$\emptyset/\neg\emptyset$	$\emptyset/\neg\emptyset$	$\emptyset/\neg\emptyset$

Figure 2-5 Matrice à neuf intersections d'Egenhofer 3x3 (Source : [Normand, 2002])

Les relations peuvent être regroupées en cinq catégories soit : disjonction, adjacence, intersection, inclusion et égalité. Ces analyses sont de types

booléens, c'est-à-dire qu'il y a, ou pas, existence d'une relation spécifique entre les objets. Contrairement aux relations métriques, les relations topologiques n'utilisent pas d'unité de mesure et elles demeurent les mêmes lorsque l'on change ou transforme le système de référence. Les analyses topologiques telles que présentées par la matrice à neuf intersections d'Egenhofer sont des analyses effectuées sur des données vectorielles. Dans le mode matriciel, la définition des objets ne permet pas d'étudier les relations qu'ils ont entre eux. Il est cependant possible d'analyser la relation de voisinage entre les cellules [Pumain et Saint-Julien, 2004]. Selon Collet [1992], le voisinage d'une cellule peut être défini comme l'ensemble des 4 cellules adjacentes à ses côtés (voisinage restreint) ou alors étendu aux 8 cellules, qui lui sont contiguës par ses côtés et ses sommets (voisinage étendu). Ces relations de voisinage dans une structure de données matricielle sont implicites par la position des cellules dans la grille. Les relations de disjonction sur une même couche matricielle ou entre couche et d'égalité entre couches sont également possibles ainsi que l'identification des relations d'intersection et d'inclusion entre couches matricielles de différentes résolutions. En matriciel, les relations topologiques sont appliquées sur les cellules de la grille et leur valeur d'attribut, elles ne permettent pas de déterminer les relations topologiques entre objets.

(4) Opérateurs d'analyse 3D

Les analyses 3D sont souvent réalisées en matriciel et sont, dans la plupart des cas, basées sur des données d'élévation. De nombreuses opérations permettent de dériver des données d'élévation, des informations comme la pente, l'orientation de la pente, son inclinaison ou encore d'effectuer des calculs de volume ou de visibilité. Les analyses 3D sont des analyses avancées, c'est-à-dire qu'elles sont réalisées à l'aide d'une combinaison d'opérateurs métriques et topologiques. Elles nécessitent des calculs métriques sur les valeurs d'élévation et nécessitent souvent des opérateurs topologiques pour la réalisation de l'analyse. Par exemple, dans une grille matricielle, la pente attribuée à une cellule dépend de l'élévation de celle-ci et de l'élévation des

cellules voisines. En matriciel, le calcul de la pente est obtenu par le passage d'un filtre sur la couche d'élévation. Ce filtre est représenté par une fenêtre glissante dont le centre passe successivement sur chaque cellule et y calcule un gradient selon la valeur des cellules voisines. Les propriétés de voisinage de la structure matricielle permettent ainsi le calcul de ces analyses.

(5) Opérateurs d'analyse de proximité

Les analyses de proximité permettent d'analyser ce qui se trouve à l'intérieur d'une certaine distance d'un objet, ou de l'ensemble des objets d'une même classe [Bédard, 2003]. Les analyses de proximités regroupent des analyses de distance simples et des analyses de distance impliquant une mesure de coût (ex. : chemin le plus court). Les analyses de distance sont réalisées à l'aide d'opérateurs métriques tandis que les analyses de coût font appel à une combinaison d'opérateurs métriques et topologiques. Une mesure de distance peut s'effectuer en ligne droite ou suivant un réseau. La construction d'un buffer autour d'un objet est un exemple d'analyse de distance en ligne droite. Ces analyses sont réalisables en vectoriel et en matriciel. La mesure de distance suivant un réseau s'effectue, quant à elle, uniquement en mode vectoriel. Une analyse de proximité impliquant une mesure de coût peut s'effectuer de son côté en mode matriciel ou suivant un réseau. En mode matriciel, ce type d'analyse résulte en une matrice de coût et permet, entre autres, d'effectuer une analyse de chemin optimal. Une analyse de coût suivant un réseau permet également la détermination d'un chemin optimal, mais s'effectue seulement en vectoriel. Les structures matricielles et vectorielles peuvent donc être utilisées pour effectuer des analyses de proximité. Toutefois, dans le cas de la structure matricielle, la résolution de la grille est un facteur qui influence énormément les mesures de distances. Ce facteur n'est donc pas à négliger lors de la réalisation d'une analyse de proximité en mode matriciel.

(6) Opérateurs de superposition spatiale

Une façon simple de créer ou d'identifier des relations spatiales est par une procédure de superposition spatiale (*Spatial Overlay*). Une analyse de superposition spatiale est réalisée à l'aide d'opérateurs topologiques, mais contrairement aux opérateurs topologiques simples servant à identifier s'il y a présence ou non de relation spatiale entre les objets, l'analyse de superposition spatiale permet de trouver les relations spatiales entre des données. Cette analyse consiste à superposer différentes données qui partagent le même territoire. La réalisation de ce type d'analyse s'effectue beaucoup plus simplement et rapidement en matriciel [Bonham-Carter, 1994; Chrisman, 1997; Demers, 2002]. Une fois les données transformées dans un même système matriciel, la superposition s'effectue très simplement par l'application d'opérateurs arithmétiques ou booléens sur les valeurs des cellules. La figure 2-6, tirée de Chrisman [Chrisman, 1997], illustre la réalisation d'une opération de superposition en mode matriciel.

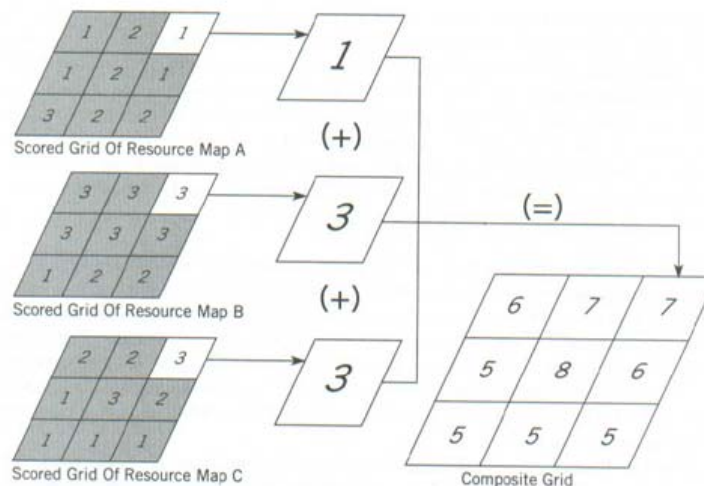


Figure 2-6 Opération de superposition spatiale en mode matriciel (Source : [Chrisman, 1997])

La superposition en mode vectoriel est beaucoup plus complexe. La superposition de données polygonales, par exemple, nécessite la création de nouveaux polygones représentant les intersections possibles. Si le nombre de

données est important, cette procédure peut devenir très complexe et peut nécessiter énormément de temps de traitement [Chrisman, 1997; Demers, 2002].

En conclusion de cette section, l'analyse spatiale est un processus très important dans l'étude des phénomènes spatiaux et comme nous venons de le voir il existe, dans les SIG, un grand nombre d'opérateurs d'analyse spatiale tout aussi différents les uns des autres. Certains s'effectuent plus aisément avec une structure de données vectorielle comme les opérateurs métriques, les opérateurs topologiques et les analyses de réseaux. Les analyses 3D, les superpositions spatiales ainsi que les analyses impliquant des relations de voisinage sont de leur côté, mieux adaptées à une structure de données matricielle. La simplicité de la structure matricielle et la simplicité des traitements effectués sur ce type de données ont largement favorisé l'émergence d'un langage communément appelé algèbre de cartes ou map algebra. Initié par Tomlin [Tomlin, 1990] vers la fin des années 80, ce langage est aujourd'hui implanté dans la plupart des SIG pour permettre l'analyse de données matricielles. Il est décrit plus en détail dans la section suivante.

2.3.3 Map Algebra

Le map algebra est un langage qui regroupe un ensemble de procédures et de fonctions élémentaires. Ce langage repose sur un modèle matriciel. Un modèle matriciel est composé de plusieurs couches, chacune d'elle représente un attribut particulier d'une zone d'étude. Les couches qui forment un modèle matriciel doivent être de même taille et de même orientation. Une couche peut représenter un attribut comme la pente, le type de sol ou la densité de population. Une opération de map algebra s'exécute par couche. C'est-à-dire que chaque opération accepte en entrée, une ou plusieurs couches et génère une nouvelle couche en sortie. En contrôlant l'ordre dans lequel les opérations sont exécutées et en utilisant le résultat d'une opération comme entrée d'une autre, il est possible de combiner les opérations pour former des procédures.

De cette manière, une foule d'analyses peuvent être effectuées simplement en combinant quelques opérations [Tomlin, 1990]. Certaines analyses sont simples et ne nécessitent qu'une seule opération tandis que d'autres plus complexes en nécessite plusieurs, comme la détermination du chemin optimal. La puissance de ce langage réside dans les fonctions qui sont exécutées par les opérations et dans la manière dont les opérations sont combinées pour former des procédures. Tomlin [Tomlin, 1990], propose une classification des fonctions élémentaires pour les données matricielles. Il définit 3 types de fonctions, (1) les fonctions locales, (2) les fonctions focales et (3) les fonctions zonales.

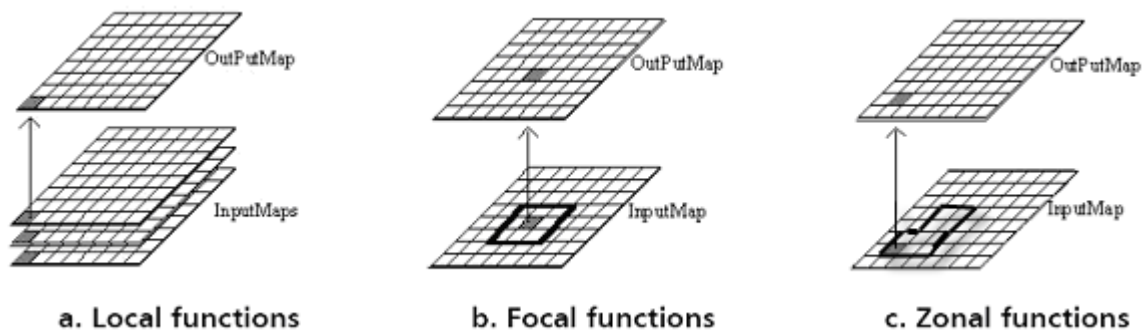


Figure 2-7 Types de fonctions de map algebra (Source : [Câmara et *al.*, 2005])

(1) Fonctions locales : Chaque nouvelle valeur des cellules de la couche résultante est le résultat d'un calcul effectué sur les valeurs existantes, d'une ou plusieurs couches, sur cette même localisation. Parmi les fonctions locales, on retrouve les analyses de superposition ou les classifications (Figure 2 -7 a)

(2) Fonctions focales : Chaque nouvelle valeur des cellules de la couche résultante est le résultat d'un calcul effectué sur les valeurs des cellules voisines. Le voisinage peut être restreint aux 8 cellules contiguës ou étendu selon une distance donnée. Une analyse de pente fait appel à une fonction focale. (Figure 2 -7 b)

(3) Fonctions zonales : Chaque nouvelle valeur des cellules de la couche résultante est le résultat d'un calcul effectué sur les valeurs des cellules voisines selon une zone donnée. Les zones peuvent être de formes et de tailles variées. Ce type de fonction permet de calculer des statistiques d'une couche d'information (ex. altitude) en fonction d'un zonage défini dans une autre couche d'information (ex. limite administrative). Une fonction zonale peut être, par exemple, le calcul de l'altitude moyenne de chacune des zones administratives. (Figure 2 -7 c)

Les fonctions sont essentiellement de simples opérateurs mathématiques, arithmétiques ou logiques appliqués sur les valeurs des cellules de manière locale, focale ou zonale [Tomlin, 1990]. Simple d'utilisation, le map algebra est devenu un langage standard pour l'analyse de données matricielles. Il permet de réaliser une foule d'analyses spatiales matricielles à l'aide de quelques fonctions simples (somme, moyenne, min, max, etc.) [Burrough et McDonnell, 1998].

En conclusion de cette section 2.3 sur la gestion de l'information à référence spatiale selon une approche transactionnelle, nous rappelons ici que les outils transactionnels tels que les SIG sont optimisés pour la saisie, le stockage, la mise à jour, la sécurité et l'intégrité des données spatiales. Tel que présenté, les SIG offrent une grande variété d'opérateurs d'analyse spatiale permettant l'exploration d'information à référence spatiale. Toutefois, la structure normalisée typiquement implantée dans les outils transactionnels implique un nombre élevé de tables et par conséquent, un grand nombre de jointures entre ces tables lors des opérations. Ceci complexifie les analyses et ralentit les temps de traitements. Typiquement, les organisations ne se servent des outils SIG que pour des données détaillées et aucune donnée d'archives n'est utilisée sauf pour le cas des bases de données spatio-temporelles. Il est donc pratiquement impossible d'effectuer des analyses temporelles et il est difficile de rendre rapidement accessibles des données agrégées. Les données ainsi produites sont de natures opérationnelles et sont complexes à exploiter à des fins décisionnelles. La section suivante présente l'approche décisionnelle,

approche orientée vers l'analyse et l'exploration de l'information à des fins décisionnelles.

2.4 Gestion et exploration de l'information à référence spatiale selon une approche décisionnelle

Les données produites par les outils transactionnels sont difficilement exploitables à des fins décisionnelles. En effet, les hauts dirigeants et les analystes éprouvent des difficultés à prendre des décisions en exploitants de telles données. Non seulement, ils sont souvent peu familiers avec les technologies informatiques, mais en plus, ils ont besoin d'informations agrégées, de comparaisons rapides dans l'espace et le temps, de synthèses sur de grandes quantités de données, de mesures de tendances ainsi que d'autres analyses complexes. Pour supporter ce type de prise de décision, il est essentiel de pouvoir réaliser les analyses de manière très rapide et très intuitive sur des données spatiales (ou non), autant détaillées qu'agrégées ainsi que sur des données historiques et actuelles. Étant donné que les outils transactionnels ne sont pas conçus pour soutenir l'analyse décisionnelle, de nouveaux outils ont été développés pour spécifiquement pallier ce besoin. Ces nouveaux outils d'analyse regroupés sous le terme d'intelligence d'affaires ou de Business Intelligence (BI) sont optimisés pour faciliter les analyses complexes et pour améliorer la performance des requêtes nécessitant plusieurs milliers d'occurrences. Les solutions de BI les plus connues et les plus utilisées sont les outils OLAP (On-Line Analytical Processing) [Codd et *al.*, 1993; Thomsen, 2002].

Dans les sections suivantes, nous présentons les principaux concepts OLAP ainsi que les concepts SOLAP (Spatial On-Line Analytical Processing) soit ceux spécifiques à la gestion de l'information à référence spatiale selon l'approche décisionnelle.

2.4.1 Concepts OLAP

Cette section ne présente qu'un aperçu des principaux concepts OLAP (On-Line Analytical Processing), étant donné qu'ils ont fait l'objet de centaines de publications depuis une quinzaine d'années. Pour plus de détails, le lecteur est invité à consulter [Codd et al., 1993; OLAP Council, 1995; Rivest, 2000; Thomsen, 2002].

Les outils OLAP se définissent comme «une nouvelle catégorie de logiciels axés sur l'exploration et l'analyse rapide des données selon une approche multidimensionnelle à plusieurs niveaux d'agrégation » [Caron, 1998].

La technologie OLAP est basée sur **l'approche multidimensionnelle**. Cette approche repose sur de nouveaux concepts tels que les dimensions, les membres, les mesures, les faits et les cubes de données [Kimball et Ross, 2002; Thomsen, 2002].

Une **dimension** est un axe ou un thème d'analyse. L'âge, le domaine d'étude et la provenance sont des exemples de dimensions. Une dimension est composée de **membres** organisés selon une hiérarchie, elle-même basée sur le niveau de détail des membres. Pour une dimension « provenance », les continents, les pays et les provinces sont des exemples de niveaux hiérarchiques et le Canada est un exemple de membre appartenant au niveau pays. Les dimensions d'un OLAP peuvent être de plusieurs types : thématique, temporelle ou spatiale (non géométrique) [Bédard et al., 2008].

Une **mesure** est un attribut souvent numérique (ex. nombre d'étudiants) qui est analysé en fonction des différentes dimensions (ex. âge, domaine d'étude, provenance). Les mesures peuvent être agrégées selon les hiérarchies des dimensions (ex. par pays et par continent), formant ainsi une vue plus globale de l'information [Rivest et al., 2001]. Un ensemble des mesures agrégées selon un ensemble de dimensions et les membres de ces dimensions forment un **cube** ou un **hypercube** [Thomsen, 2002]. À l'intérieur de ce cube, les différentes combinaisons de membres de dimensions et de mesures forment

les **faits**. Un fait représente ainsi la valeur d'une mesure et les membres de chacune des dimensions ayant servi à son calcul (ex. 30 étudiants en géomatique âgés de 19 ans). Les faits peuvent être de types détaillés ou agrégés [Kimball et Ross, 2002]. Les faits détaillés forment le niveau de détail le plus fin d'analyse, ils sont composés des mesures calculées et des membres les plus fins de chacune des dimensions ayant servi à calculer les mesures. Parfois, ils correspondent aux données de bases issues directement des bases de données sources, mais souvent ils représentent déjà un agrégat relativement à ces dernières. Quant aux faits agrégés, ils sont le résultat des agrégations effectuées sur les faits détaillés. Une agrégation est un traitement qui consiste généralement à l'utilisation d'opérateurs statistiques comme la somme, la moyenne, le maximum et le minimum effectué sur des données détaillées permettant de les regrouper pour former des données agrégées. Dans un cube de données, l'ensemble des combinaisons de membres de dimensions et de mesures est typiquement **précalculé** afin d'améliorer la performance, en terme de vitesse d'exécution, des analyses.

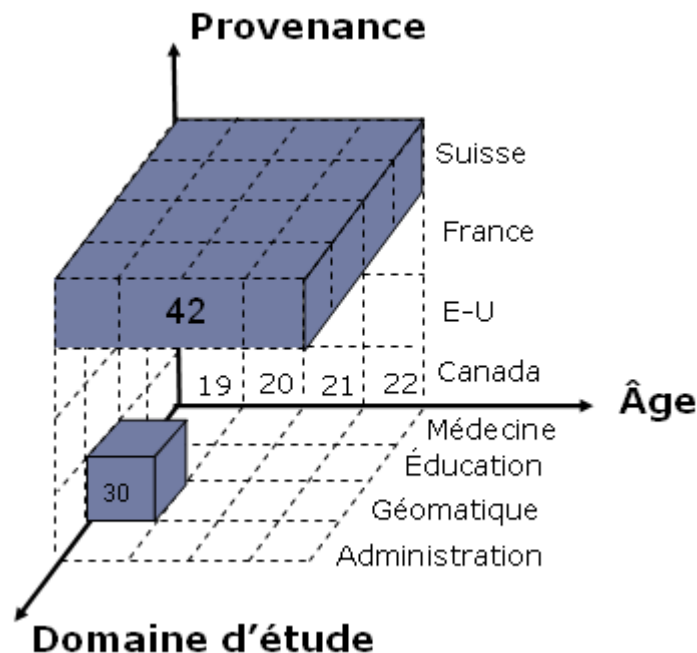


Figure 2-8 Exemple d'un cube de données

La figure 2-8 illustre un cube de données formé de trois dimensions (domaine d'étude, âge et provenance). La mesure analysée en fonction de ces trois dimensions est le nombre d'étudiants. Le fait, représenté par un petit cube, indique que pour le secteur de la géomatique il y a 30 étudiants canadiens âgés de 19 ans. La strate du haut représente l'agrégation du nombre d'étudiants (42) originaires de la Suisse, âgés entre 19 et 22 ans pour chacun des domaines d'études.

L'architecture générale d'un système OLAP se compose, en général, de trois modules : une base de données multidimensionnelle, un serveur OLAP et le client OLAP. Le client OLAP permet aux utilisateurs d'explorer et d'analyser les données à l'aide de différentes méthodes de visualisation et de simples opérateurs adaptés à la navigation dans les bases de données multidimensionnelles [Bédard et *al.*, 1997]. Les principaux opérateurs OLAP sont le forage, le remontage, le forage latéral et le pivot. Ils permettent de naviguer et d'explorer très facilement et intuitivement dans les données en utilisant de simples cliques de souris. Il n'est donc pas nécessaire pour l'utilisateur de maîtriser un langage d'interrogation comme SQL pour accéder aux données situées dans la base de données multidimensionnelle [Marchand et *al.*, 2003]. En interagissant directement avec les données, l'utilisateur peut, ainsi, se concentrer sur les résultats de ses analyses plutôt que sur les procédures nécessaires à leur réalisation. Les outils OLAP sont optimisés dans le but de permettre une analyse facile et rapide des données. La majorité des combinaisons possibles d'analyse sont précalculées pour accélérer les temps de réponse et les données sont présentées par thèmes (dimensions) et par niveaux de détails (hiérarchies). Ainsi, ces outils sont idéaux pour l'exécution de requêtes globales ou agrégatives. Les temps d'exécution des analyses sont toujours minimisés, tant au niveau détaillé qu'agrégé, ce qui permet à l'utilisateur de conserver son fil de pensée.

Dans les outils OLAP, l'information à référence spatiale est gérée seulement de manière nominale. En effet, même si les données sont composées d'une composante spatiale, celle-ci n'est exploitée qu'en utilisant les noms de lieu. Il

fut démontré par Caron [Caron, 1998] que les outils OLAP possédaient un potentiel pour l'analyse spatio-temporelle, par contre, lorsque la composante spatiale est traitée de manière nominale, ce potentiel est grandement limité. En effet, sans vision cartographique, il est difficile de pouvoir observer la distribution spatio-temporelle des phénomènes analysés et de détecter des regroupements spatiaux (ex. le long du St-Laurent) différents des regroupements prédéfinis dans la dimension spatiale (ex. MRC). En ce sens, des équipes de recherche ont proposé une approche multidimensionnelle pour la gestion et l'analyse de données à référence spatiale. Les équipes de Bédard et *al.* [Bédard et *al.*, 1997] et Han et *al.* [Han et *al.*, 1998] furent parmi les premières à proposer une telle approche. Ces travaux ont conduit à la création d'une nouvelle catégorie d'outils, les outils SOLAP (Spatial On-Line Analytical Processing). Les concepts SOLAP sont approfondis à la section suivante.

2.4.2 Concepts SOLAP

L'approche multidimensionnelle présentée à la section 2.4.1 est également à la base des outils SOLAP. Toutefois, afin de tenir compte des particularités associées aux données spatiales, cette approche a été enrichie de concepts spatiaux. Dans les bases de données multidimensionnelles à référence spatiale, l'information spatiale peut s'intégrer sous forme de dimensions ou de mesures. Dans cette section, nous abordons tout d'abord (1) la notion de dimension spatiale et (2) de mesure spatiale. Nous décrivons ensuite (3) la notion de cubes spatiaux pour terminer avec (4) le concept de dimension d'opérateurs topologiques spatio-temporels.

(1) Dimensions spatiales

Dans la littérature, on distingue trois types de dimensions spatiales : (a) les dimensions spatiales non-géométriques, (b) les dimensions spatiales géométriques et (c) les dimensions spatiales géométriques mixtes [Han et *al.*, 1998; Bédard et *al.*, 2001, 2008; Rivest et *al.*, 2003]. La figure 2-9 illustre ces trois types de dimensions spatiales.

(a) Dimension spatiale non-géométrique : Dimension dont les membres ne possèdent pas de primitive géométrique. La référence spatiale est dite nominale. Ce type de dimension ne permet pas de représentation cartographique et peut être implanté comme une dimension descriptive. Ce type de dimension spatiale est le seul exploitable dans les outils OLAP conventionnels.

(b) Dimension spatiale géométrique : Dimension dont les membres de tous les niveaux possèdent une primitive géométrique.

(c) Dimension spatiale géométrique mixte : Dimension dont les membres ne possèdent pas tous une géométrie, les autres membres ayant une référence nominale seulement. Au minimum, il doit y avoir un membre représenté géométriquement et un autre représenté nominalement.

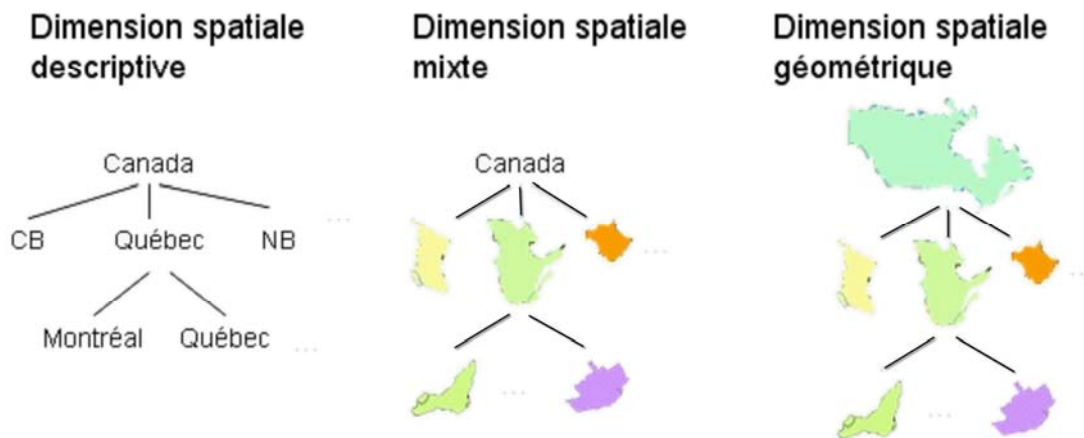


Figure 2-9 Trois types de dimensions spatiales présentement supportés par le SOLAP (Source : [Rivest et *al.*, 2003]).

Comme la structure vectorielle est la seule utilisée jusqu'ici dans les applications SOLAP (excepté dans les travaux de Miquel et *al.*, [Miquel et *al.*, 2002]), la structure des primitives géométriques utilisée dans les dimensions spatiales (hormis non-géométrique) fait référence à cette structure vectorielle.

(2) Mesures spatiales

En plus des mesures descriptives numériques, on distingue maintenant trois types de mesures spatiales : (a) les mesures spatiales numériques, (b) les mesures spatiales géométriques et (c) les mesures totalement spatiales [Bédard et *al.*, 2008].

(a) Mesure spatiale numérique : Mesure correspondant au résultat d'une analyse spatiale. Le résultat est numérique et peut représenter une superficie, une distance ou un nombre de voisins, par exemple. Ce type de mesure est également appelé mesure spatiale non-géométrique.

(b) Mesure spatiale géométrique : Mesure correspondant à un ensemble de coordonnées ou de pointeurs vers des objets géométriques. Ces objets sont obtenus par application d'une opération géométrique (ex. union, intersection, concaténation, etc.) sur les membres d'une ou plusieurs dimensions spatiales.

(c) Mesure totalement spatiale : Mesure correspondant à la combinaison d'une valeur numérique et de sa géométrie associée. Par exemple, la superficie d'un incendie de forêt et sa géométrie.

(3) Cubes spatiaux

Les cubes spatiaux sont des cubes pour lesquels des membres de dimensions ou des faits (via des mesures spatiales) sont référencés spatialement et peuvent être représentés sur des cartes [Bédard et *al.*, 2008]. Suite à ces présents travaux de recherche, nous distinguons maintenant deux types de cubes spatiaux, les cubes d'objets et les cubes matriciels. Nous verrons à la section 3-3 la définition de ces deux types de cubes spatiaux.

(4) Dimension d'opérateurs topologiques spatio-temporels

En plus des dimensions et des mesures spatiales, une base de données multidimensionnelle à référence spatiale peut être enrichie d'une dimension d'opérateurs topologiques spatio-temporels. Tel que mentionné à la section 1-

1, ce type de dimension initié par les travaux de Marchand [Marchand, 2004] a permis d'améliorer les outils SOLAP en termes d'analyse spatiale topologique. Nommée Spatio-Temporal Topological Operator Dimension (STTOD), cette dimension implante une hiérarchie d'opérateurs topologiques spatio-temporels permettant de décrire les relations existantes entre les membres des dimensions spatiales. Cette hiérarchie couvre les 3 domaines possibles des relations topologiques spatio-temporelles : spatial, temporel et spatio-temporel. Des opérateurs globaux forment le niveau supérieur et se raffinent pour former le niveau détaillé de la hiérarchie. La figure 2-10 présente une vue simplifiée de la hiérarchie. Le membre « same time, same place » forme le niveau supérieur et la figure présente des raffinements possibles de ce membre lorsqu'une opération de forage est appliquée sur celui-ci.

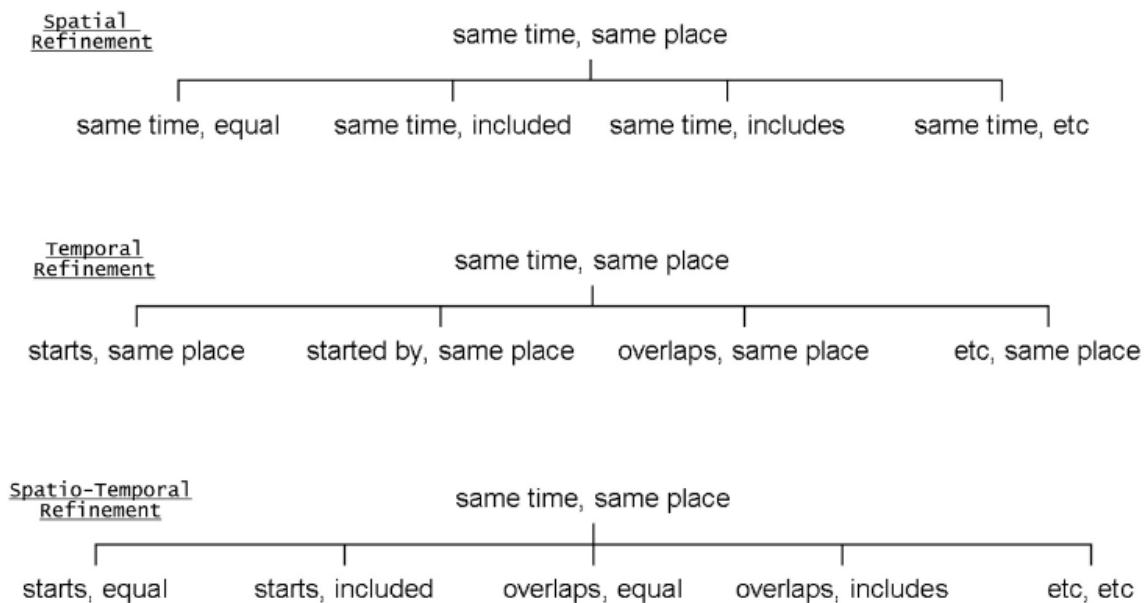


Figure 2-10 Raffinements possible du membre « Same time, Same place » (Source : [Marchand, 2004])

L'apport de cette recherche est considérable puisqu'une telle dimension permet d'exploiter des analyses topologiques dans un SOLAP. Comme l'ensemble des relations spatio-temporelles sont décrites au préalable, les temps de

traitements sont compatibles. Cependant, le précalcul et le stockage de l'ensemble de ces relations représentent un effort à ne pas négliger

Les outils SOLAP détiennent tous les avantages des outils OLAP et permettent en plus l'analyse et la gestion de données spatiales. Ils possèdent également des opérateurs de navigation pour explorer l'ensemble des données spatiales au moyen d'une vue cartographique. Les principaux opérateurs spatiaux de navigation sont le forage spatial, le remontage spatial et le forage latéral spatial. Ils permettent de naviguer d'un niveau, ou d'un membre, à l'autre dans une dimension spatiale.

L'ajout de la composante cartographique aux outils OLAP conventionnels présente d'importants avantages pour l'analyse de phénomènes géographiques et spatio-temporels. La représentation cartographique facilite l'extraction d'informations qui permettent de mieux comprendre les phénomènes analysés. Ces informations peuvent porter sur les caractéristiques spatiales des phénomènes (position, forme, orientation, etc.), sur les relations spatiales entre phénomènes (adjacence, inclusion, proximité, etc.) ou encore sur leur distribution (régulière, aléatoire, concentrée, etc.) [Bédard et *al.*, 2007]. Cependant, outre l'utilisation d'une dimension d'opérateurs topologiques spatio-temporels, les fonctionnalités d'analyse spatiale sont actuellement limitées dans ces outils. Cette limite est principalement due à l'utilisation de la structure de données vectorielle. La seule structure de données présentement supportée par les technologies existantes sur le marché est la structure vectorielle [Proulx et Rivest, 2007]. L'utilisation unique de cette structure de données limite les possibilités d'analyses notamment en ce qui concerne l'analyse de phénomènes continus.

2.5 Conclusion

Ce chapitre a dans un premier temps présenté des concepts généraux de l'information à référence spatiale. Les types de phénomènes géographiques et leur représentation sous forme d'objet ou de couverture ont été définis. La

notion de couverture, servant généralement à la représentation de phénomènes continus, a fait l'objet d'une description plus approfondie. Les types de modèles conceptuels de la réalité et les structures de données géométriques vectorielles et matricielles ont ensuite été présentés.

La seconde section de ce chapitre a présenté les notions plus spécifiques à la gestion et l'exploitation de l'information à référence spatiale dans les outils transactionnels. Ces outils supportent les structures de données vectorielles et matricielles et ils offrent de nombreuses fonctionnalités d'analyse spatiale.

Finalement, la dernière section de ce chapitre présente les notions relatives à la gestion et l'exploitation de l'information à référence spatiale au sein des outils décisionnels. Les concepts généraux OLAP et ceux de l'approche multidimensionnelle ont d'abord été exposés. Ensuite une description des concepts spécifiques à l'exploitation de données spatiales au sein de bases de données multidimensionnelles a été fournie. Ceci nous a permis de mettre en évidence le fait que, dans les outils de type SOLAP, la gestion de l'information spatiale s'effectue uniquement à l'aide d'une structure de données vectorielle ce qui amoindrit le pouvoir d'analyse des données spatiales. Le chapitre qui suit propose des concepts permettant l'intégration d'une structure de données matricielle dans les cubes de données spatiaux.

CHAPITRE 3 CONCEPTS THÉORIQUES PERMETTANT L'INTÉGRATION DU MATRICIEL DANS LES CUBES SPATIAUX

3.1 Introduction

Le chapitre précédent a introduit les notions de base de cette recherche, notamment sur la gestion et l'exploration de l'information à référence spatiale dans les systèmes transactionnels et décisionnels. Nous avons constaté que dans les systèmes décisionnels de type SOLAP, la seule structure de données présentement utilisée est la structure de données vectorielle et qu'aucune approche n'a encore été développée pour concevoir et implanter des cubes de données supportant le matriciel. C'est ainsi que ce présent chapitre propose un cadre théorique permettant l'intégration de données matricielles dans les cubes spatiaux. Deux approches sont alors définies : la première approche propose d'intégrer le matriciel dans les dimensions spatiales et la seconde propose de l'intégrer dans les faits. Ces deux approches conduisent aux concepts théoriques développés dans le cadre de cette recherche.

3.2 Intégration du matriciel dans les dimensions spatiales

Tel que présenté à la section 2.4.2, on distingue actuellement dans la littérature trois types de dimensions spatiales (descriptives, géométriques et mixtes) [Han et al, 1998; Bédard et al, 2001, 2008; Rivest et al., 2003]. Dans ces dimensions (hormis descriptive), la structure des primitives géométriques utilisée est vectorielle. Si on considère la possibilité d'exploiter une structure matricielle dans les dimensions spatiales en plus de la structure vectorielle déjà utilisée, on obtient sept types de dimensions spatiales (dont quatre nouveaux

types) tels qu'illustrés à la figure 3-1. Pour mieux distinguer les différents types, nous avons renommé la dimension spatiale géométrique actuelle en dimension spatiale géométrique vectorielle et la dimension spatiale mixte en dimension spatiale mixte vectorielle. De plus, le terme « hybride » a été introduit afin d'identifier les dimensions combinant le vectoriel et le matriciel et ne doit pas être confondu avec la notion de « mixité » qui réfère à la nature descriptive et géométrique des données.

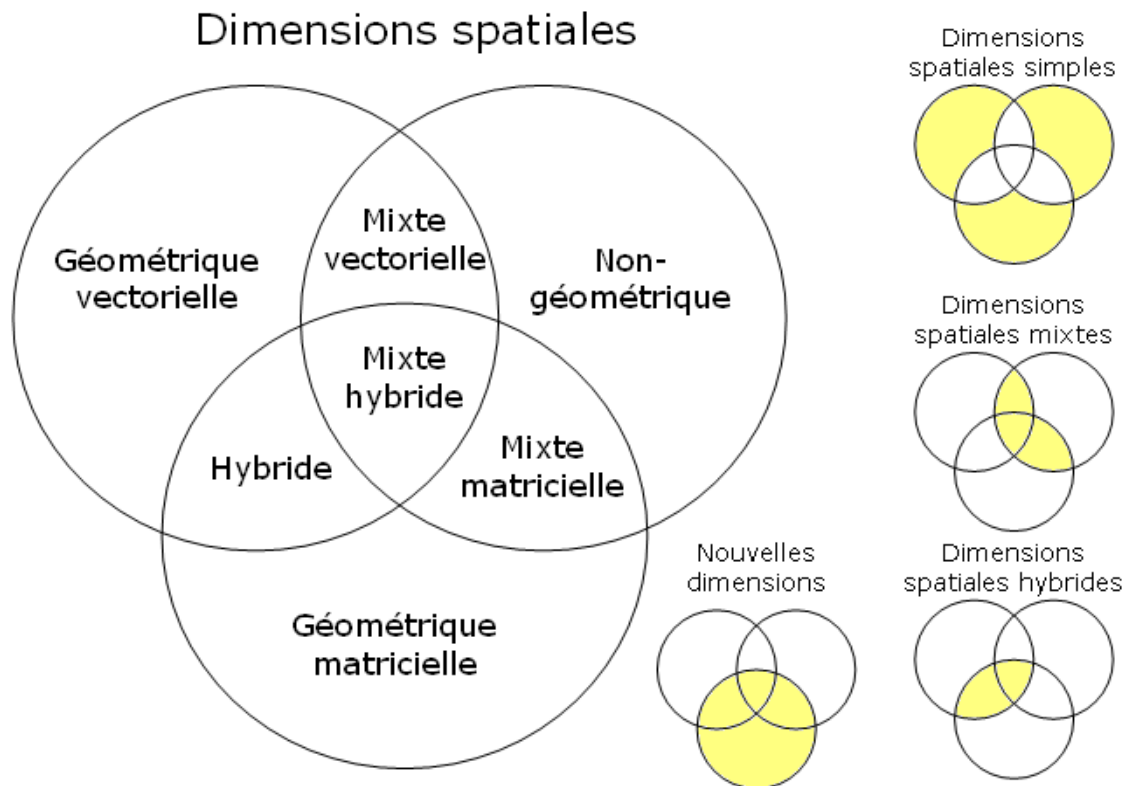


Figure 3-1 Les types de dimensions spatiales (Source : [Bédard et *al.*, 2008])

L'intégration de la structure matricielle dans les dimensions spatiales a ainsi nécessité la définition de quatre nouveaux types de dimensions spatiales : (1) la dimension géométrique matricielle, (2) la dimension géométrique hybride, (3) la dimension mixte matricielle et (4) la dimension mixte hybride.

(1) Dimension géométrique matricielle : Une dimension géométrique matricielle est constituée d'au moins un niveau matriciel et si le nombre de

niveaux est supérieur à 1, différentes résolutions géométriques sont employées pour former la hiérarchie. La plus grande résolution est utilisée pour le niveau le plus détaillé et les résolutions agrégées sont des multiples entiers de la résolution du niveau détaillé. Un membre correspond à une cellule. Les membres d'un même niveau hiérarchique sont donc de même taille et de même forme. Une dimension géométrique matricielle permet une représentation matricielle du territoire à différentes échelles. La figure suivante illustre ce type de dimension spatiale.

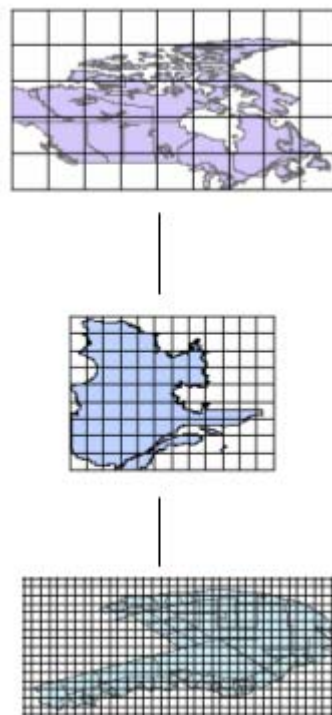


Figure 3-2 Dimension géométrique matricielle

(2) Dimension géométrique hybride : Une dimension géométrique hybride est constituée d'au moins un niveau matriciel et d'au moins un niveau vectoriel et ne doit pas contenir de niveau non-géométrique. Ces niveaux peuvent être dans n'importe quel ordre et lorsque le nombre de niveaux hiérarchiques est supérieur à deux, toute combinaison vectoriel – matriciel est possible. Ce type de dimension permet, par exemple, d'analyser un phénomène de manière globale selon un découpage vectoriel prédéfini comme celui des pays et des

provinces et d'ensuite raffiner l'analyse à un niveau fin matriciel sans limite particulière. Elle permet également d'analyser un phénomène de manière globale selon un découpage matriciel et de raffiner l'analyse vers un découpage fin vectoriel. La figure suivante illustre deux dimensions géométriques hybrides, l'une composée d'un niveau fin matriciel et l'autre d'un niveau fin vectoriel.

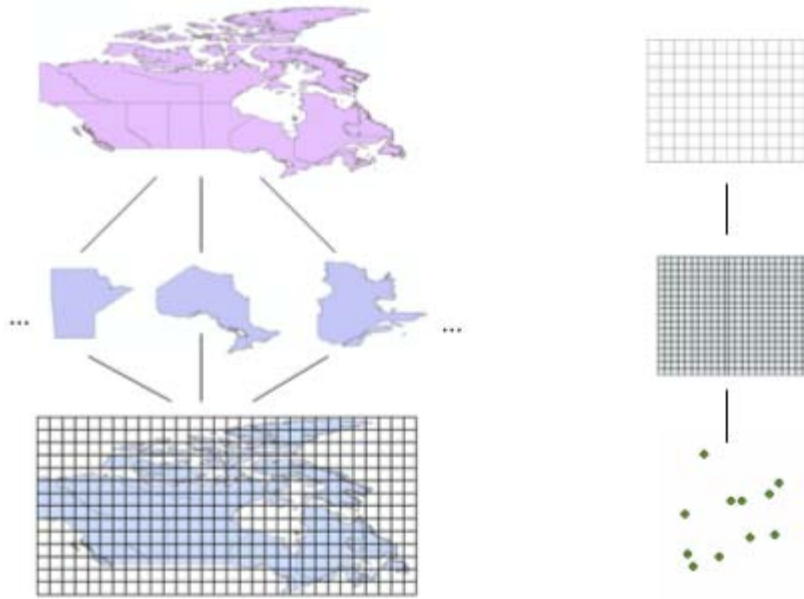


Figure 3-3 Dimension géométrique hybride

(3) Dimension mixte matricielle : Une dimension mixte matricielle est constituée d'au moins un niveau matriciel et d'au moins un niveau descriptif et ne doit pas contenir de niveau vectoriel. Ces niveaux peuvent être dans n'importe quel ordre et lorsque le nombre de niveaux hiérarchiques est supérieur à deux, toute combinaison matriciel - descriptif est possible. Ce type de dimension permet, par exemple, d'analyser un phénomène de manière globale selon un découpage dont on ne possède pas la géométrie et d'ensuite raffiner l'analyse à un niveau fin matriciel sans limite particulière. La figure suivante illustre ce type de dimension.

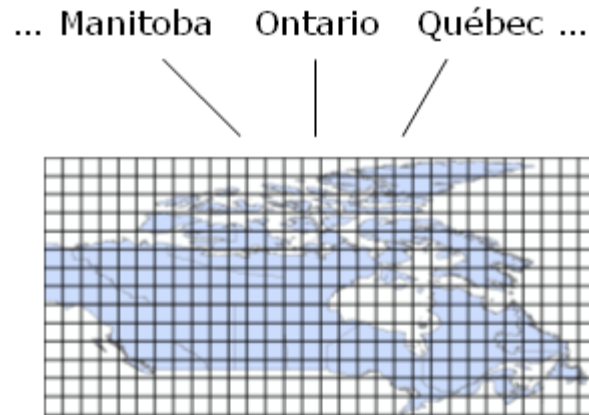


Figure 3-4 Dimension mixte matricielle

(4) Dimension mixte hybride : Une dimension mixte hybride est constituée d'au moins trois niveaux. Elle est composée d'au moins un niveau descriptif, d'au moins un niveau matriciel et d'au moins un niveau vectoriel, et ceci, dans n'importe quel ordre. Ce type de dimension permet de combiner tout type d'information géographique, par exemple de l'information nominale sur les pays au niveau agrégé, suivi d'un découpage vectoriel des provinces canadiennes et finalement d'une représentation matricielle au niveau détaillé. La figure suivante illustre ce type de dimension.

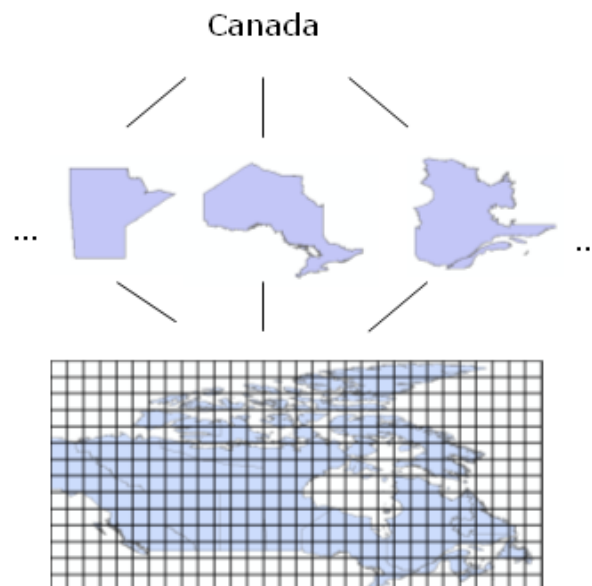


Figure 3-5 Dimension mixte hybride

Nous devons noter ici que les dimensions géométriques hybrides et les dimensions mixtes hybrides constituées consécutivement d'un niveau matriciel et d'un niveau vectoriel polygonal ou linéaire sont fondamentalement des dimensions dites « non strictes », c'est-à-dire qu'elles impliquent inévitablement des relations multiples entre les niveaux. Ceci est dû au fait que les niveaux ne s'emboîtent pas parfaitement, c'est-à-dire que les membres du niveau supérieur ne correspondent pas à un nombre entier de membres de niveau inférieur. Si on regarde par exemple la relation entre un niveau fin matriciel et le découpage des pays, un pays peut être constitué de plusieurs cellules et les cellules qui chevauchent les frontières appartiennent à plus d'un pays (figure 3-6).

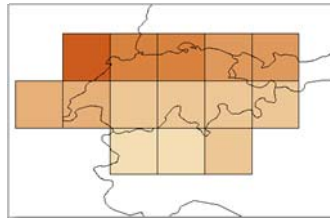


Figure 3-6 Relation entre la Suisse et les cellules qui la compose (Source : [Proulx et *al.*, 2008])

Ce type de relation n'est pas simple à gérer dans un contexte multidimensionnel et nécessite des règles d'implantation particulières. Il faut établir des règles de passage d'un niveau à l'autre qui peuvent tenir compte par exemple du degré de recouvrement entre un membre de niveau supérieur et les membres du niveau inférieur. Nous verrons au chapitre 4 différentes méthodes pour implanter ce type de relation.

De plus, un cube peut être composé d'une ou de plusieurs dimensions spatiales et ces dimensions peuvent être de n'importe quel type. Il est donc possible de retrouver dans un même cube une dimension géométrique vectorielle et une dimension géométrique matricielle. Toutes combinaisons de dimensions sont théoriquement possibles, tout dépend des phénomènes analysés et de la manière dont l'utilisateur désire les représenter. Par exemple, un utilisateur

peut vouloir représenter un phénomène de population de manière continue sur le territoire. Il peut vouloir représenter ce même phénomène, mais selon un découpage administratif prédéfini ou encore, le représenter à tour de rôle de manière continue et discrète. L'utilisation d'une dimension géométrique matricielle permettrait la représentation et l'analyse de ce phénomène de manière continue. Une dimension géométrique vectorielle de son côté pourrait être composée d'une hiérarchie « monde – continents – pays » et ainsi permettre la représentation et l'analyse de ce même phénomène, mais de manière discrète. La combinaison des deux dimensions, vectorielle et matricielle, dans un même cube permettrait à l'utilisateur de représenter à la fois le phénomène de manière continue et discrète. Finalement, l'utilisation d'une dimension hybride permettrait la représentation du phénomène de manière continue et discrète dépendamment du niveau hiérarchique utilisé. Il serait également possible d'utiliser le concept de hiérarchie alternative, c'est-à-dire de combiner dans une même dimension deux hiérarchies et de les utiliser à tour de rôle. Comme nous le verrons au chapitre 5, suivant la structure de données utilisée dans les dimensions spatiales, le potentiel de l'utilisateur à exécuter des analyses spatiales ainsi que la manière de représenter et d'analyser les phénomènes changent significativement.

3.3 Intégration du matriciel dans les faits

Par l'utilisation de dimensions spatiales, nous avons vu dans la section précédente une première approche pour intégrer le matriciel dans les cubes de données. Cette présente section propose une deuxième approche, qui implique l'intégration du matriciel dans les faits. L'intégration du matriciel dans les faits correspond à utiliser la cellule d'une grille matricielle comme élément de base pour l'analyse. C'est-à-dire que chaque cellule d'une grille matricielle correspond à un fait détaillé dans un cube de données (l'intersection des membres détaillés de chaque dimension). La résolution de la grille matricielle joue ainsi un rôle très important puisque plus la résolution est fine, plus le nombre de faits détaillés augmente. Ces faits peuvent ensuite être agrégés

selon les hiérarchies des dimensions présentes dans le cube pour former les faits agrégés. Les faits agrégés correspondent ainsi à des agrégats de cellules.

Nous avons initié cette approche en effectuant une analogie entre un modèle matriciel et un cube de données (Figure 3-7).

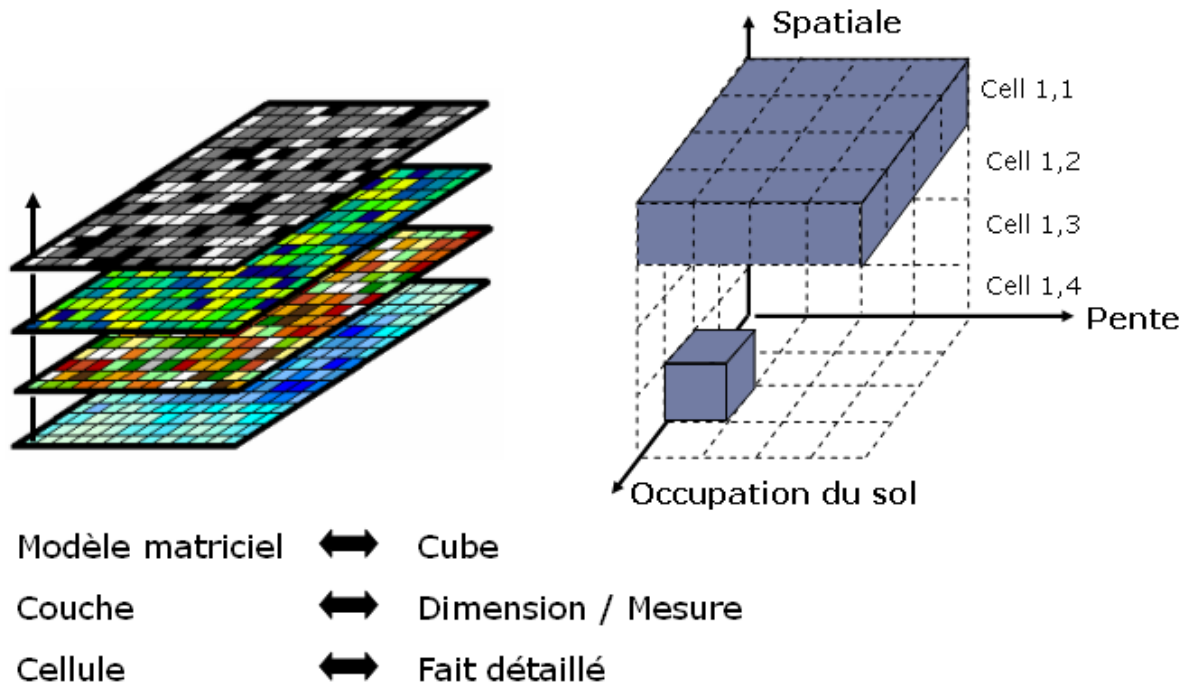


Figure 3-7 Analogie entre un modèle matriciel et un cube de données

Dans un modèle matriciel, les éléments de base servant à l'analyse sont les cellules qui forment la grille matricielle, alors que dans un cube de données, ces éléments de base correspondent aux faits détaillés. Chacune des cellules d'un modèle matriciel peut ainsi correspondre à un fait détaillé dans un cube de données. Dans un modèle matriciel, la coordonnée d'une cellule correspond au croisement entre la dimension ligne et la dimension colonne, tandis que dans un cube, la coordonnée d'une cellule correspond au croisement entre les différentes dimensions du cube. Dans un modèle matriciel, les différentes valeurs des cellules sont représentées par des couches (ex. type de sol, température, densité de population) alors que dans un cube de données, les attributs des éléments analysés sont contenus dans les membres des

dimensions ou dans les mesures. Les valeurs des cellules d'une couche matricielle peuvent ainsi correspondre soit aux membres d'une dimension soit à une mesure d'un cube de données dépendamment de l'information représentée (numérique ou textuelle) et de l'analyse désirée. Dans un cube, les informations textuelles sont généralement contenues dans les membres des dimensions tandis qu'on retrouve généralement l'information numérique sous forme de mesures [Kimball et Ross, 2002]. Les informations contenues dans les dimensions sont utilisées comme éléments de sélection dans l'analyse tandis que les mesures sont analysées en fonction des différentes dimensions et sont représentées sous forme de carte, tableau ou graphique.

Cette approche permet de représenter une structure de données matricielle sous une forme multidimensionnelle. Elle permet de naviguer dans un cube de données de la même manière dont on navigue dans une couche matricielle, c'est-à-dire cellule par cellule ou fait par fait. Il sera ainsi possible de bénéficier des opérateurs d'analyse matricielle dans un cube de données en appliquant les opérateurs sur les mesures des faits (cf. section 4.2.3), ce qui permettra d'enrichir le potentiel d'analyse spatiale dans les outils SOLAP. Tout comme un modèle matriciel, cette approche a l'avantage de pouvoir représenter tout type d'objets (point, ligne, polygone) dans un même cube. Il est également simple d'analyser les corrélations entre différents phénomènes (ex. densité de la population, élévation, température) pourvu qu'ils puissent être représentés à l'aide d'une même structure matricielle. Le cube matriciel et le modèle matriciel sont des concepts très semblables. Toutefois, le cube matriciel se distingue par la structure hiérarchique de ses dimensions. Cette structure permet d'agréger l'information selon les hiérarchies présentes dans le cube. Les cellules peuvent ainsi être agrégées pour former des agrégats de cellules (faits agrégés).

Nous avons vu dans la section précédente que l'intégration de la structure matricielle dans les dimensions conduisait à la définition de 4 nouveaux types de dimensions spatiales. L'intégration du matriciel dans les faits conduit de son côté à la définition de deux nouveaux types de cubes spatiaux. Les sections

suivantes présentent les définitions de ces cubes ainsi que l'approche permettant de choisir le bon type de cube.

3.3.1 Définition des types de cubes spatiaux

Afin de différencier les types de cubes uniquement vectoriels (c.-à-d. ceux développés jusqu'à présent excepté pour les travaux de Miquel et *al.*, [Miquel et *al.*, 2002]) de ceux utilisant le matriciel dans les faits, nous avons défini deux types de cubes spatiaux : les **cubes d'objets** et les **cubes matriciels** (Figure 3-8). La distinction entre les deux types de cubes s'effectue au niveau des faits détaillés utilisés. Les faits détaillés d'un cube d'objets sont des objets ou des phénomènes discrets (ex. personne, magasin) tandis que les faits détaillés d'un cube matriciel sont les cellules d'une grille matricielle, les éléments réguliers d'un territoire continu.

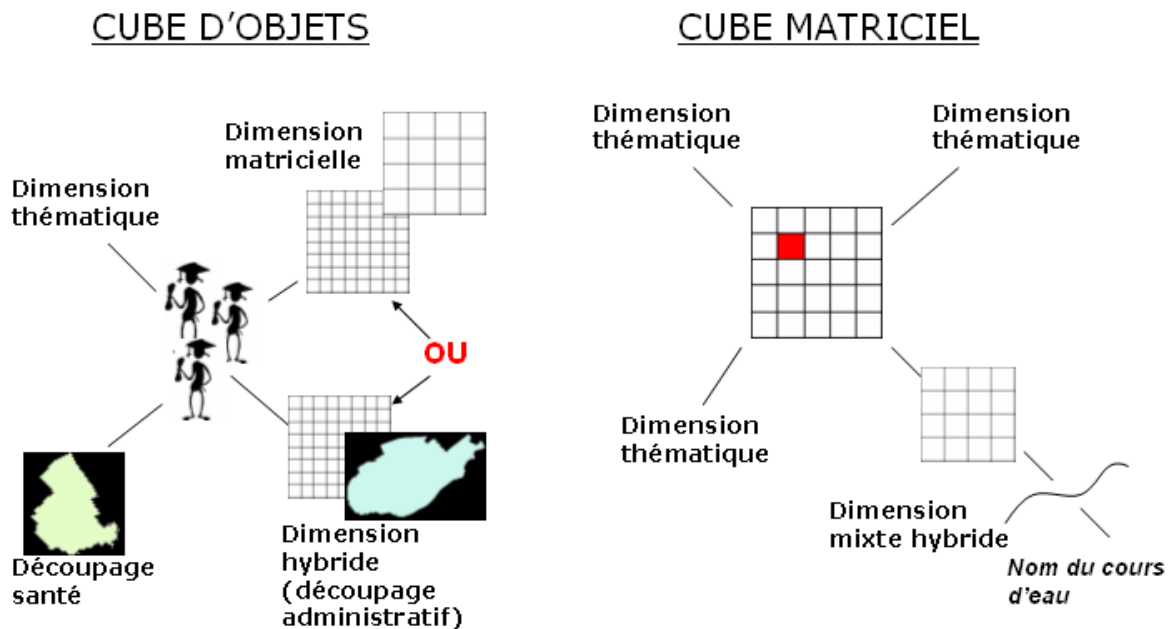


Figure 3-8 Types de cubes spatiaux

Tel qu'il a été présenté à la section 2.4.2, les cubes spatiaux sont des cubes pour lesquels des membres de dimensions ou des faits sont référencés

spatialement, nous définissons ainsi les deux types de cubes spatiaux comme suit :

Cube d'objets : Un cube d'objets est un cube dont les faits détaillés sont des objets discrets possédant ou non une géométrie. Dans le cas où les faits ne possèdent pas de géométrie, le cube doit contenir au moins une dimension dont certains ou tous les membres sont géométriques. Dans le cas où les faits possèdent une géométrie, cette géométrie peut être propre au fait (pouvant être dérivée ou non des dimensions) ou correspondre à la géométrie d'un membre spatial.

Cube matriciel : Un cube matriciel est un cube dont les faits détaillés sont les cellules d'une grille matricielle. Pour le cube matriciel, les faits ont toujours une géométrie et cette géométrie peut être propre au fait ou correspondre à la géométrie d'un membre spatial.

Toutes combinaisons de types de cubes et de dimensions spatiales sont théoriquement possibles. Un cube d'objets peut ainsi contenir une dimension matricielle et un cube matriciel une dimension vectorielle.

Remarquons ici que l'utilisation d'une dimension vectorielle dans un cube matriciel ou l'utilisation d'une dimension matricielle dans un cube d'objets peut introduire des relations multiples entre les faits et la dimension. Tout comme les dimensions « non strictes », les relations multiples entre les faits et les membres d'une dimension ne sont pas simples à gérer dans un contexte multidimensionnel et nécessitent des règles d'implantation. Différentes méthodes pour gérer ce type de relations seront présentées au chapitre 4.

3.3.2 Choix du type de cube

Dans les systèmes transactionnels, le choix de la structure de données dépend en grande partie de la modélisation conceptuelle qu'on se fait de la réalité. Une modélisation par objets résulte en une utilisation d'une structure vectorielle et une modélisation par couverture en une utilisation d'une structure matricielle.

Dans les systèmes décisionnels, le choix du type de cube peut s'effectuer sensiblement de la même manière. Toutefois, comme les cubes de données sont construits dans un contexte précis, et sont orientés pour l'analyse nous parlerons plutôt d'**analyse par objets** et d'**analyse par couverture**. L'analyse par objets est simplement une expression utilisée pour faire référence au type d'analyse effectué sur des données modélisées selon un modèle par objets et l'analyse par couverture correspond au type d'analyse effectué sur des données modélisées selon un modèle par couverture (cf. section 2.2.3). Ainsi, le choix du type de cube dépend en grande partie du type d'analyse réalisé.

Une **analyse par objets** a pour but d'étudier des objets discrets en fonction de leurs caractéristiques, de leur position ou des relations qui existent avec d'autres objets (ex. analyser des unités d'habitation en fonction de leur valeur marchande, de leur type et de leur position sur le territoire). Dans ce contexte, le cube d'objets est le plus approprié. Les faits détaillés étant des objets discrets (unités d'habitation) et les dimensions du cube des caractéristiques de ces objets (valeur marchande, type d'unité d'habitation, quartier).

Une **analyse par couverture** a pour but d'étudier un territoire particulier en termes de variations spatiales de caractéristiques, et ce, sans limites particulières (ex. analyse de la variation spatiale de la valeur marchande en fonction de la distance à un centre d'intérêt). L'utilisation d'un cube pour effectuer une analyse par couverture implique des dimensions contenant des caractéristiques reliées au territoire, les faits détaillés sont ainsi forcément des éléments de ce territoire. Dans ce contexte, l'utilisation d'un cube matriciel est tout à fait appropriée, les faits détaillés étant des cellules d'une grille matricielle et les dimensions des caractéristiques de ces cellules (valeur marchande de la cellule qui peut être la moyenne des valeurs marchandes des unités d'habitation incluses dans la cellule et la distance à un centre d'intérêt). Conséquemment, **une analyse par objets conduit à la construction d'un cube d'objets et une analyse par couverture à la construction d'un cube matriciel.**

3.4 Conclusion

Nous avons, dans ce chapitre, défini un cadre théorique permettant l'intégration de données matricielles au sein de cubes de données spatiaux. Nous avons proposé deux approches complémentaires, la première implique l'intégration du matriciel dans les dimensions spatiales et la seconde dans les faits. L'intégration du matriciel dans les dimensions a nécessité la définition de 4 nouveaux types de dimensions spatiales : les dimensions géométriques matricielles, les dimensions hybrides, les dimensions mixtes matricielles et les dimensions mixtes hybrides. L'intégration du matriciel dans les faits a, de son côté, conduit à la définition de deux types de cubes spatiaux : le cube d'objets et le cube matriciel. Ces deux types de cubes se distinguent au niveau des faits détaillés utilisés. Les faits détaillés d'un cube d'objets sont des objets ou des phénomènes discrets, tandis que les faits détaillés d'un cube matriciel sont les cellules d'une grille matricielle. Un cube matriciel est en réalité une modélisation multidimensionnelle d'un modèle matriciel et étant donné que les faits détaillés de ce type de cube sont des cellules matricielles, il sera éventuellement possible d'effectuer des opérations d'analyses matricielles sur les faits comme on le fait dans un modèle matriciel.

Les deux approches proposées dans ce chapitre permettent de concevoir et d'implanter des cubes avec des données matricielles. Dans un cube d'objets, l'intégration du matriciel s'effectue uniquement au niveau des dimensions spatiales et le type d'analyse effectué avec ce type de cube est une analyse par objets. Tandis que dans un cube matriciel l'intégration de données matricielles s'effectue à la fois dans les faits et dans les dimensions spatiales, il permet d'analyser les données d'une tout autre manière, jusqu'ici non exploitée dans les cubes, en permettant une analyse des données par couverture. Les concepts théoriques présentés dans ce chapitre sont les résultats théoriques proposés par cette recherche, mais ils ont tout d'abord été testés et raffinés lors d'expérimentations exploratoires. Une fois les concepts testés et raffinés, ils ont ensuite été utilisés dans un projet impliquant un

partenaire de la chaire de recherche industrielle en bases de données géospatiales décisionnelles. Ce dernier projet a permis de démontrer l'utilité de nos concepts et leur potentiel à répondre à un besoin actuel. Le chapitre suivant décrit en détail ces expérimentations.

CHAPITRE 4 EXPÉRIMENTATIONS

4.1 Introduction

Les nouveaux concepts permettant l'intégration de la structure matricielle dans les cubes spatiaux ont été présentés au chapitre précédent. Ces nouveaux concepts ont été testés dans le cadre d'expérimentations exploratoires et ensuite validés dans le cadre d'une expérimentation formelle. Nous présentons dans ce chapitre ces différentes expérimentations. Il est important de mentionner que l'enjeu de ces expérimentations n'est pas le choix entre l'utilisation d'une structure vectorielle ou matricielle, mais bien de démontrer qu'il est possible d'utiliser le matriciel dans un cube de données et de tester les différentes approches proposées au chapitre précédent.

L'outil JMap Spatial OLAP, développé par l'équipe de la chaire de recherche CRSNG en bases de données géospatiales décisionnelles et commercialisé par KHEOPS Technologies, a été utilisé pour la réalisation des expérimentations (hormis l'expérimentation 3). L'outil JMap Spatial OLAP ne supporte actuellement pas de structure matricielle en tant que telle pour les analyses multidimensionnelles et l'implantation d'une structure matricielle supportée par cet outil dépassait largement le cadre de cette recherche. Par conséquent, les expérimentations ont été réalisées à l'aide d'une « simulation matricielle » plutôt qu'une structure matricielle pure. La simulation matricielle représente le territoire de la même manière qu'une structure pure, mais chaque cellule est un polygone carré issu d'une vectorisation des cellules d'une grille matricielle. La « simulation matricielle » apporte l'avantage d'avoir un lien direct avec un outil SOLAP existant et de permettre de tester rapidement et concrètement les concepts développés. Les opérateurs de navigation (forage, remontage, forage latéral) peuvent être utilisés sur les cellules (dans le cas où elles sont des membres d'une dimension) étant donné qu'elles sont en réalité des polygones

vectoriels. L'utilisation d'une structure vectorielle pour représenter les données sous la forme de cellules permet de tester les concepts élaborés tout en étant supportée par l'application JMap Spatial OLAP.

4.2 Expérimentations exploratoires

Trois expérimentations exploratoires ont permis de tester et de raffiner les nouveaux concepts théoriques définis au chapitre précédent. Ces expérimentations ne sont pas exhaustives, c'est-à-dire que nous n'avons pas testé l'ensemble des combinaisons possibles entre les types de dimensions spatiales et les types de cubes spatiaux. Cependant, les expérimentations réalisées sont suffisantes pour démontrer la faisabilité des nouveaux concepts. Ces expérimentations ont été réalisées à l'aide de données réelles et fictives et les jeux de données ont été choisis en fonction des concepts à tester. L'expérimentation 1 a été réalisée à l'aide des données du projet Muscamags et elle a permis de tester l'intégration du matriciel dans les dimensions spatiales d'un cube d'objets. L'expérimentation 2 a été réalisée à l'aide des données du cours GMT-20853 Travaux pratiques en SIG et télédétection du département des sciences géomatiques de l'Université Laval et elle a permis de tester l'intégration du matriciel dans les faits et dans les dimensions spatiales d'un cube. La troisième et dernière expérimentation exploratoire réalisée à l'aide des données des inventaires forestiers de la forêt Montmorency a, quant à elle, permis d'explorer le potentiel d'analyse spatiale matricielle à la volée dans un cube matriciel.

4.2.1 Expérimentation exploratoire 1

La première expérimentation a été réalisée à l'aide des données du projet Muscamags (Multi-Scale Multi-Agent Geo-Simulation) (<http://www2.ift.ulaval.ca/~MUSCAMAGS/ProjectDetailedDescription.htm>). Ce projet, financé par le Réseau GEOIDE, implique 6 chercheurs de 4 universités du Canada (U. Laval, U. McMaster, U. Queen's et U. Wilfrid Laurier) ainsi que

plusieurs partenaires industriels et gouvernementaux. Le projet a pour but de développer une plate-forme de simulations qui permettra d'aider les décideurs à développer des stratégies d'interventions en cas de crise (ex. panne de courant affectant toute la ville), de comparer différents scénarios envisagés ainsi que d'anticiper les conséquences de ces scénarios et interventions. La région ciblée dans le cadre de ce projet est la grande région de Québec. À partir de données représentant les principales caractéristiques d'un échantillon représentatif de la population de Québec, l'équipe a recréé la ville de Québec dans un environnement synthétique ainsi que les agents représentant ses citoyens.

Les données

Les données utilisées dans le cadre de notre expérimentation sont celles sur les unités d'habitation, les ménages et les principaux services aux citoyens de la Communauté métropolitaine de Québec (CMQ). On y observe la distribution de ces entités sur l'ensemble du territoire couvert, les endroits à forte densité de population, les endroits moins habités, les endroits bien desservis par certains services, etc. Les données utilisées proviennent de plusieurs sources, elles ont été fournies et traitées par Céline Drolet, candidate à la maîtrise au Département d'informatique de l'Université Laval. Voici une description de ces données :

- **Unités d'habitation** : les données sur les unités d'habitation proviennent du rôle d'évaluation foncière de 2005. Nous avons l'information sur le type d'unité d'habitation : appartement, unité multiple ou unité simple.
- **Ménages** : Les données sur les ménages proviennent de l'enquête O-D 2001 [gouvernement du Québec, 2007] (les données de cette enquête ont été fournies par Marius Thériault professeur au Département d'aménagement du territoire et de développement régional de l'Université Laval). L'enquête O-D est une enquête origine-destination sur les habitudes de déplacements des gens. Elle

représente un échantillon de la population de la CMQ (10%). Des facteurs d'expansions ont ensuite été appliqués à ces données pour représenter l'ensemble des ménages. Ces facteurs sont déterminés en fonction de critères statistiques bien précis pour permettre une représentation le plus juste possible de l'ensemble de la population. Les informations exploitables sont le type de ménage, les ménages avec ou sans enfant et le statut (monoparental ou en couple) pour les ménages avec enfants.

- **Services** : Les données sur les services proviennent de la Base de données Zipcom 2001. Les informations exploitables portent sur le type de service (magasin, épicerie, restaurant, loisir, éducation, santé, etc.) et sur leur localisation (isolé, sur une rue commerciale ou dans un centre commercial).

Les données ont été traitées et agrégées sur un découpage régulier matriciel ayant une résolution de 500m et ont été stockées dans une base de données Oracle 10g. Pour chaque cellule, nous avons le nombre d'entités de chaque type inclus dans la cellule et les coordonnées des points de maximum densité à l'intérieur de la cellule. Il y a un point de maximum densité pour chaque type d'information par cellule (ex, l'endroit de maximum densité des ménages sans enfant dans une cellule). La résolution de 500m a été déterminée par les responsables du projet Muscamags pour des raisons de confidentialité. Les données utilisées dans le cadre de ce projet sont des données portant sur les habitudes de déplacements des gens. Elles contiennent des informations sensibles qui permettraient potentiellement d'identifier les individus, c'est pourquoi les données ont été agrégées sur une grille minimum de 500m de côté.

4.2.1.1 Concepts testés

Le type d'analyse réalisé dans ce contexte-ci est une analyse par objets, c'est-à-dire qu'on désire analyser des objets discrets comme les unités d'habitation, les ménages et les services en fonction de caractéristiques qui leur sont

propres (type d'unité d'habitation, de ménage ou de service) et de leur localisation sur le territoire. Le cube approprié dans ce contexte est un cube d'objets et les faits détaillés sont représentés par les points de maximum densité. Dans un cube d'objets, l'intégration du matriciel est possible seulement par l'utilisation de dimensions spatiales. **Nous avons ainsi testé dans cette expérimentation l'intégration du matriciel dans les dimensions spatiales d'un cube d'objets.** Nous avons intégré une dimension géométrique matricielle nommée « Découpage régulier » composée de 3 niveaux hiérarchiques (500m, 1000m et 2000m) et une dimension géométrique hybride nommée « Découpage administratif » composée d'un niveau fin vectoriel représentant les points de maximum densité, suivi d'un niveau matriciel d'une résolution de 500m et d'un niveau vectoriel représentant les arrondissements de la CMQ. Ces dimensions sont représentées dans la figure 4-1.

La **dimension géométrique matricielle** permet d'agréger l'information analysée selon un découpage régulier sans restriction de limites administratives ou de découpages prédéfinis. Les niveaux de la hiérarchie sont composés de cellules matricielles de différentes résolutions. Cette dimension permet de visualiser l'ensemble du territoire en cellules matricielles de différentes grosseurs et d'analyser la distribution des données couvrant ces cellules. Elle permet ainsi de représenter les informations de manière continue sur l'ensemble du territoire couvert et de potentiellement faire d'autres regroupements que ceux prédéfinis par les hiérarchies (regroupements ad hoc).

La **dimension géométrique hybride** permet, de son côté, d'analyser la distribution des ménages, des unités d'habitation et des principaux services selon un découpage administratif représenté par les anciennes municipalités de la CMQ (avant les fusions). Ce découpage contient 32 arrondissements d'une superficie moyenne de 100 km². Ce qui implique que l'information sur ce découpage est représentée de manière très agrégée. Le niveau cellule (500m) permet ainsi de forer à l'intérieur d'un arrondissement et de voir la distribution

des entités selon un découpage régulier et plus fin. Finalement, le niveau fin représente les points de densité maximale à l'intérieur des cellules.

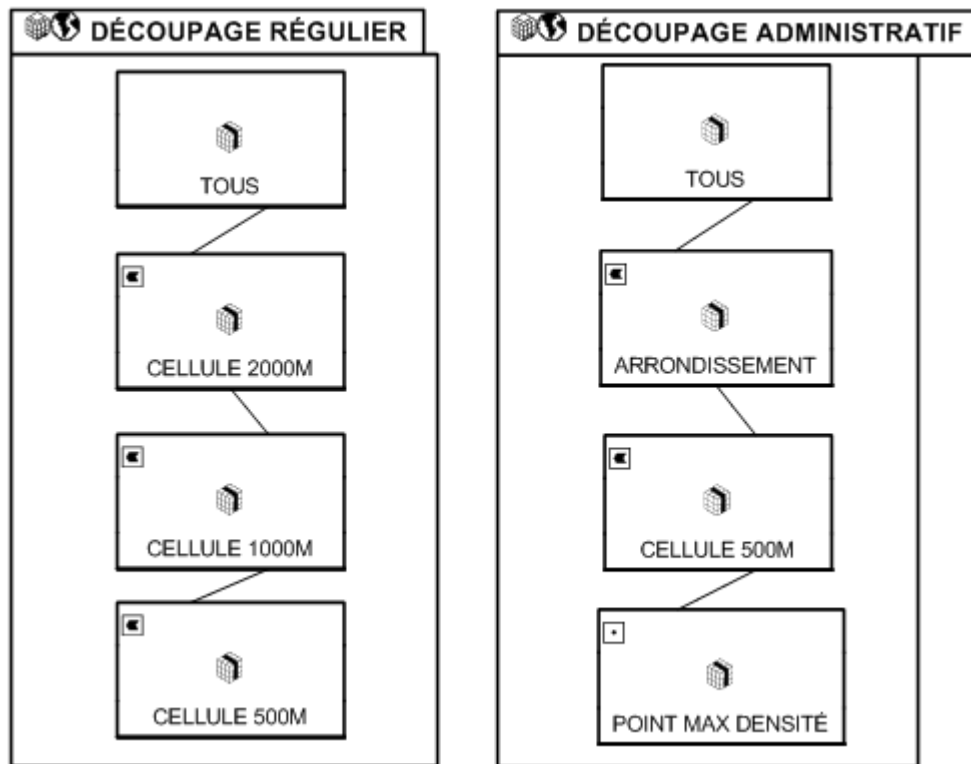


Figure 4-1 Dimension spatiale géométrique matricielle et dimension spatiale géométrique hybride

4.2.1.2 Conception du cube

Comme présenté précédemment, le cube contient une dimension géométrique matricielle et une dimension géométrique hybride. En plus de ces 2 dimensions, il est composé de 7 dimensions thématiques et de 6 mesures. Le schéma du cube est présenté à la figure 4-2 et vous trouverez en Annexe A le schéma en étoile ainsi que les diagrammes de membres des dimensions.

Les dimensions « Type de service », « Localisation des services », « Type d'unité d'habitation » et « Type de ménage » sont des caractéristiques des entités analysées. En plus de permettre une analyse de l'information de manière continue selon un découpage régulier, l'intégration du matriciel dans

les dimensions du cube a également permis d'introduire des analyses matricielles de distances.

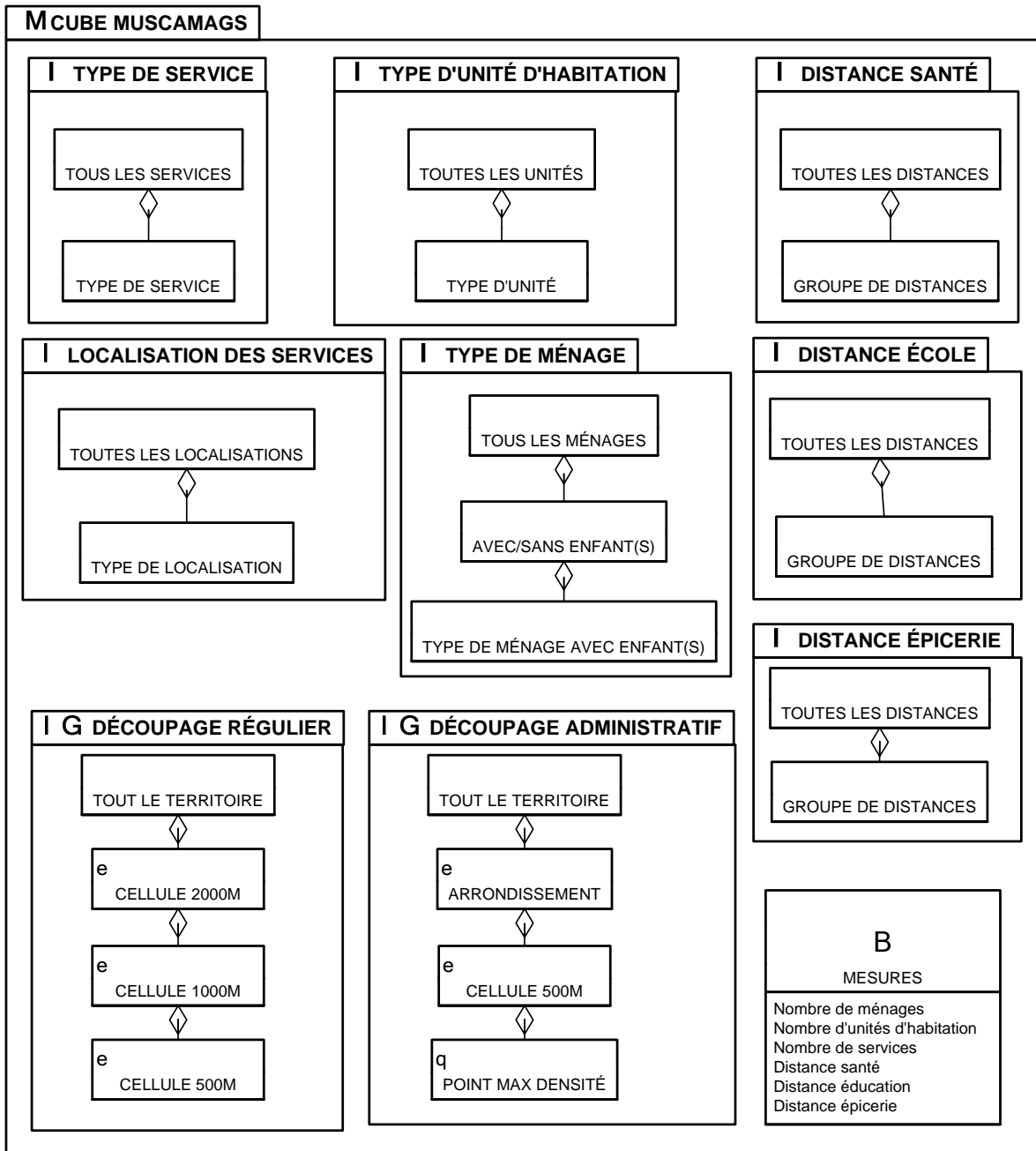


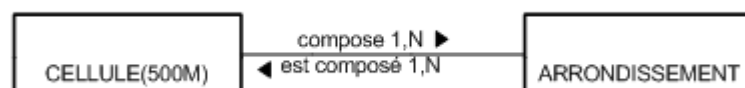
Figure 4-2 Schéma du cube Muscamags

Il est ainsi possible d'assigner aux faits détaillés des caractéristiques des cellules dans lesquelles ils se trouvent. Nous avons ainsi réalisé des analyses

matricielles de distances en ligne droite à partir des services de santé, des épiceries et des écoles. Ces analyses ont été précalculées à l'aide du logiciel ArcGIS d'ESRI version 9.2. Les résultats ont ensuite été implantés dans le cube en dimensions et en mesures. Les dimensions « Distance santé », « Distance école » et « Distance épicerie » contiennent une classification des distances résultantes tandis que les mesures du même nom contiennent les distances réelles. Ces mesures sont des mesures spatiales numériques puisqu'elles correspondent aux résultats d'opérations spatiales (cf. section 2.4.2). Les analyses de distances permettent d'évaluer l'accessibilité aux services essentiels et d'évaluer les endroits moins bien desservis par ces services. Ces analyses sont également réalisables en vectoriel, elles sont d'ailleurs plus précises, cependant elles auraient nécessité beaucoup plus de calculs. En effet, pour chaque fait détaillé représenté par les points de maximum densité, il aurait fallu déterminer le plus proche voisin et ensuite mesurer la distance entre les deux, et ce, pour les 18 235 faits détaillés et pour les 3 mesures de distance (santé, école, épicerie) ce qui aurait nécessité de mesurer 54 705 distances. Dans notre contexte, nous ne cherchions pas la précision absolue, mais plutôt un ordre de grandeur, l'analyse en matriciel était donc appropriée et elle était beaucoup plus simple et rapide à réaliser.

4.2.1.3 Limites, difficultés rencontrées et solutions apportées

Une première difficulté rencontrée réside dans l'implantation de la dimension géométrique hybride « Découpage administratif ». Une relation multiple existe dans cette dimension entre les cellules (500m) et les arrondissements. Le regroupement des cellules en arrondissements crée une relation N-N (Figure 4-3), lorsque les cellules se situent sur les frontières des arrondissements.



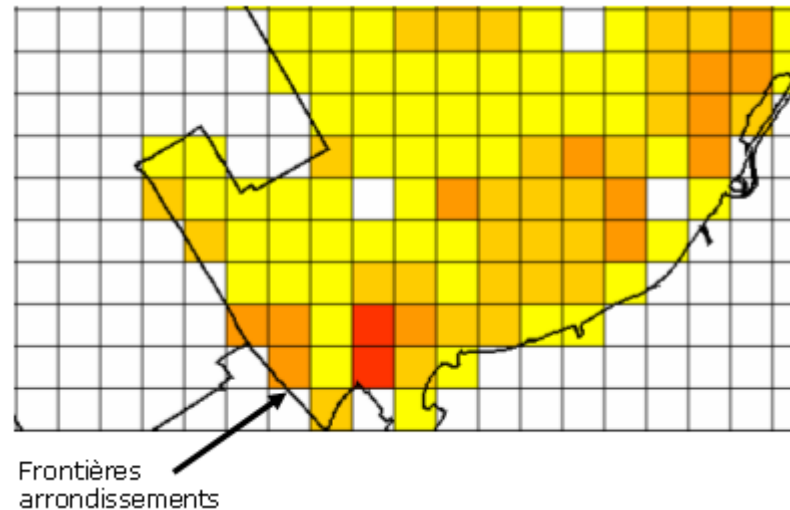


Figure 4-3 Relation N-N entre cellule et arrondissement

Afin d'être bien implanté, ce type de relation nécessite des règles d'implantations particulières. Nous décrivons ici quelques-unes des solutions envisageables pour contourner le problème.

Solution 1 : Kimball [Kimball et Ross, 2002] propose de créer une table-pont pour décrire la relation multiple. Cette solution est simple, mais le serveur utilisé doit gérer ce type de table, ce qui n'est pas le cas actuellement pour JMap Spatial OLAP.

Solution 2 : Planter la relation comme une 1-N en associant une cellule à un seul arrondissement. C'est-à-dire associer une cellule au membre de niveau supérieur qui offre le plus grand recouvrement de superficie ou qui est affecté par un attribut considéré significatif dans l'analyse. Pour l'agrégation des données, il est possible de considérer un facteur de recouvrement. Bien que cette solution soit très simple, elle peut entraîner des pertes d'informations et introduire des erreurs dans le calcul des agrégations. Dans de tels cas, l'utilisateur pourrait être averti au moyen d'un système de mises en garde [Lévesque et *al.*, 2007].

Solution 3 : Exploiter la redondance d'information dans la dimension, c'est-à-dire multiplier les membres du niveau cellule qui intersectent plus d'un

arrondissement. Cette solution nécessite la pré agrégation des données dans le cube et l'utilisation de formules considérant les doublons. Selon la mesure à agréger, il est possible d'utiliser soit un pourcentage de la superficie intersectée, soit un simple facteur qui considère le nombre de doublons. Cette solution n'entraîne aucune perte d'information. Toutefois, si le nombre de relations multiples est élevé, elle peut devenir complexe. Il faut également faire attention puisqu'il est possible d'introduire des relations N-N entre la table de faits et la dimension.

Dans le cadre de cette expérimentation, nous avons choisi de tester les solutions 2 et 3.

Pour la solution 2, nous avons simplement associé les cellules qui chevauchent plus d'un arrondissement à un seul d'entre eux. Nous les avons associées aux arrondissements qui offrent le plus grand recouvrement de superficie. Les outils de rasterisation des SIG permettent d'effectuer cette association très simplement.

La solution 3 a, pour sa part, nécessité plus de traitements. Avec l'outil JMap Spatial OLAP, lorsqu'on multiplie les membres spatiaux d'une dimension on doit aussi créer autant de géométries dans le fichier spatial. L'identifiant qui fait le lien entre la table de faits et la dimension doit être le même que celui qui fait le lien vers le fichier géométrique. Conséquemment, 2 membres ne peuvent pas être liés à la même géométrie. L'utilisation de cette solution a ainsi nécessité la création de 1 250 nouveaux membres dans la dimension et autant de géométries dans le fichier géométrique de la dimension géométrique hybride. La procédure peut s'effectuer automatiquement avec l'opérateur de *SpatialJoin* du logiciel ArcGIS d'ESRI version 9.2. En créant un spatial join ONE TO MANY entre les cellules et les arrondissements, les cellules qui intersectent plus d'un arrondissement sont multipliées, ainsi que leur géométrie. Le niveau détaillé de la dimension étant ponctuel, nous n'avons pas eu à utiliser de facteur dans les agrégations. Nous avons simplement associé les points aux

cellules s'agrégeant dans l'arrondissement dans lequel le point est situé (voir exemple de la figure 4-4).

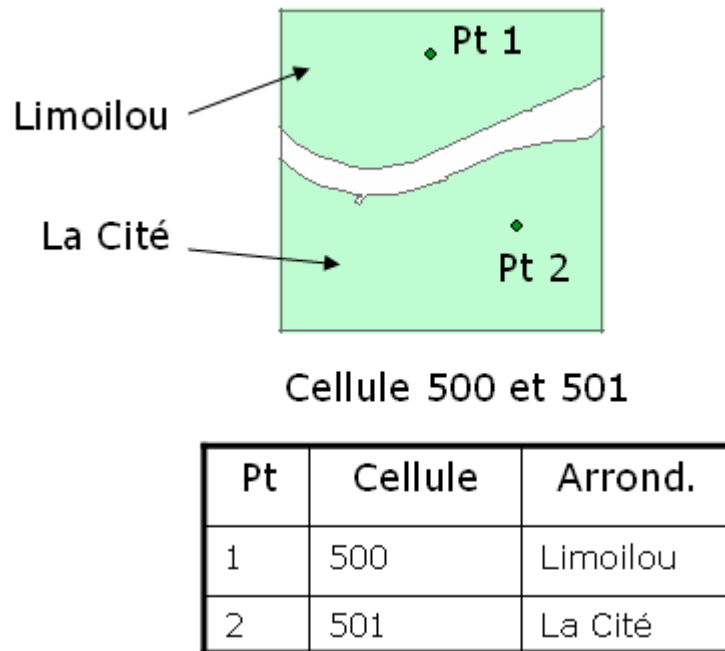


Figure 4-4 Association des points aux cellules et aux arrondissements dans la dimension géométrique hybride (découpage administratif)

4.2.1.4 Exemples d'analyses

Le cube créé permet de réaliser plusieurs analyses. En effet, il est par exemple possible de premièrement identifier le nombre de ménages avec enfant(s) par arrondissement. Ensuite, un forage peut être effectué sur un arrondissement particulier pour visualiser la localisation de ces ménages au niveau cellule et ainsi découvrir d'autres regroupements que ceux prédéfinis par les arrondissements (Figure 4-5). Comme l'illustre la figure 4-6, à l'aide des mesures de distances, il est également possible d'analyser la distance des ménages avec enfant(s) à une école et de visualiser la distribution de ces ménages sur une carte selon un découpage régulier.

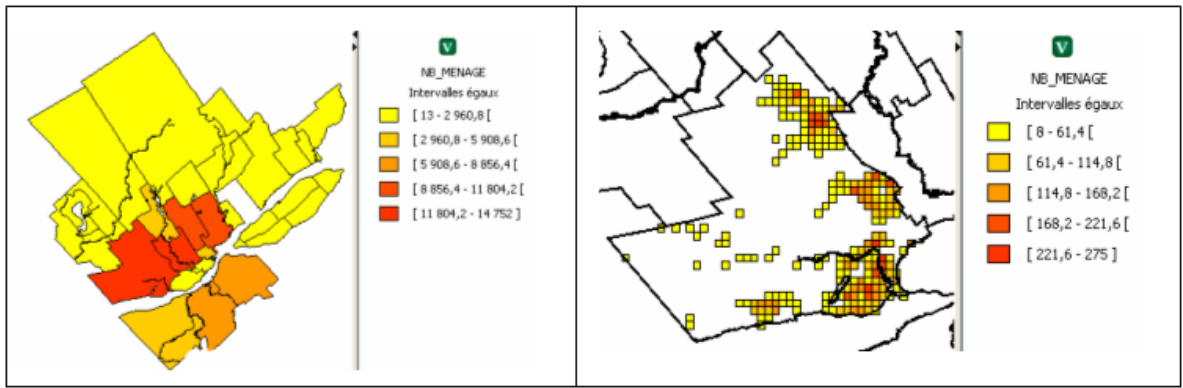


Figure 4-5 Nombre de ménages avec enfant(s) par arrondissement et forage au niveau cellule (500m)

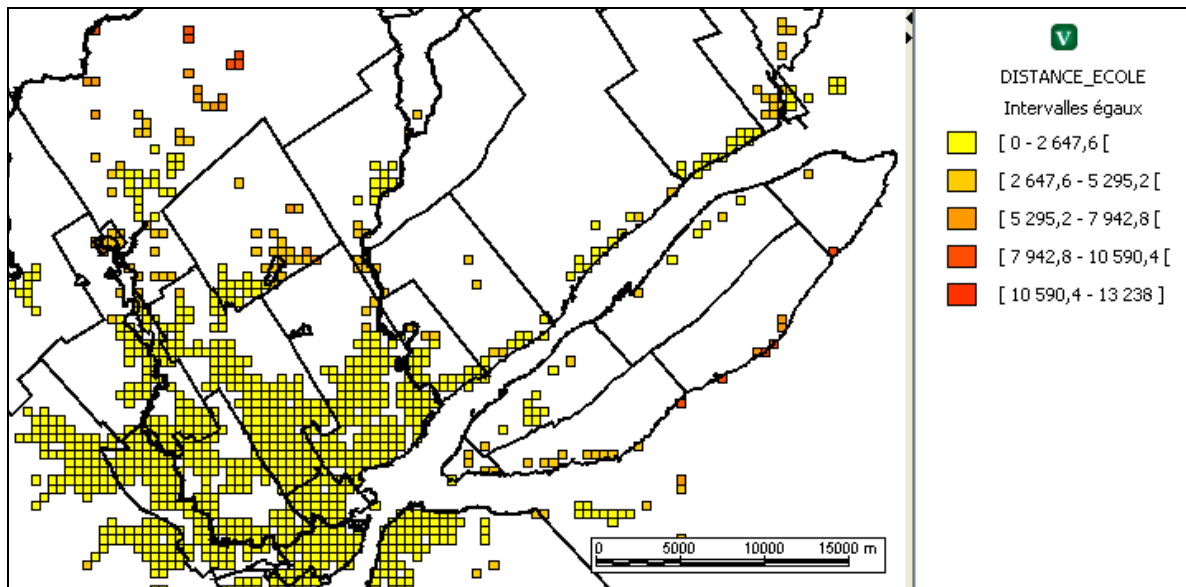


Figure 4-6 Distance des ménages avec enfant(s) à une école

4.2.2 Expérimentation exploratoire 2

La deuxième expérimentation a été réalisée à l'aide des données utilisées dans le cadre du cours GMT- 20853 Travaux pratiques en SIG et télédétection. Les données exploitées proviennent de la BNDT (Base Nationale des Données Topographiques), du zonage agricole de la CPTAQ (Commission de Protection du Territoire Agricole du Québec) et d'une Ortho-image Landsat 7. Leur

utilisation permet d'identifier et d'évaluer des tracés potentiels pour l'établissement d'un train à grande vitesse (TGV) dans le corridor ferroviaire Québec-Windsor. Les tracés doivent respecter un ensemble de critères techniques, environnementaux, économiques et sociaux, qui sont présentés en Annexe B. Trois tracés potentiels ont été proposés et sont ensuite comparés en fonction de critères comme la longueur totale, le coût global de la construction, les coûts par classe d'occupation du sol, les coûts par classe de pente, le pourcentage de bonnes terres agricoles touchées, le nombre de coupures avec des infrastructures existantes, etc.

4.2.2.1 Concepts testés

L'analyse réalisée dans cette expérimentation est une analyse par couverture dans le sens où on analyse un territoire particulier en fonction des critères à respecter, pour l'établissement d'un tracé de TGV. Ces critères à respecter visent par exemple à minimiser la pente, à minimiser les coupures avec les infrastructures existantes, à minimiser les bonnes terres agricoles, à éviter les zones urbaines, etc. On évalue les coûts potentiels à chaque endroit du territoire pour ensuite déterminer le tracé optimal, c'est-à-dire celui de moindre coût en fonction des contraintes imposées. Le cube approprié dans ce contexte est un **cube matriciel** et les faits détaillés sont des cellules d'une grille matricielle auxquelles on associe des critères et des coûts permettant la comparaison des tracés potentiels de TGV. Le cube contient une dimension spatiale géométrique hybride (Figure 4-7) composée d'un niveau fin matriciel de même résolution que les faits détaillés du cube et d'un niveau agrégé vectoriel correspondant aux tracés potentiels de TGV. Ces tracés sont les résultats d'analyses de coûts et d'analyses de tracés optimaux que nous avons précalculés et implantés dans la dimension spatiale. **Nous avons ainsi testé dans cette expérimentation l'utilisation du matriciel dans les faits et dans les dimensions spatiales. Nous avons construit un cube matriciel composé d'une dimension spatiale géométrique hybride.** Cette dimension spatiale permet, au niveau agrégé, d'analyser, de comparer et de visualiser les tracés potentiels de TGV dans leur ensemble. Le niveau fin

matriciel permet ensuite de segmenter ces tracés et d'identifier les segments affectés par certaines caractéristiques telles que la pente, l'occupation du sol et la présence d'infrastructure.

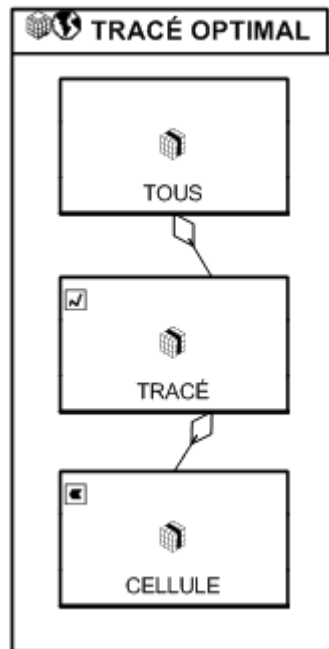


Figure 4-7 Dimension spatiale géométrique hybride

4.2.2.2 Conception du cube

Le cube construit dans cette expérimentation est un cube matriciel. Ses faits détaillés sont donc des cellules d'une grille matricielle alors que les dimensions sont des caractéristiques de ces cellules. La résolution de la grille utilisée est de 60m, ce qui donne 16 730 faits détaillés. Le choix de cette résolution résulte du fait qu'elle correspond à la fois à l'emprise d'un tracé de TGV et à un multiple entier de la résolution de l'image Landsat 7 utilisée (30m), facilitant ainsi les analyses et minimisant les traitements. En plus de la dimension spatiale géométrique hybride présentée précédemment, le cube est composé de 2 dimensions thématiques, d'une dimension spatiale descriptive et de 6 mesures. Les données descriptives sont stockées dans une base de données Oracle 10g et les données géométriques sont stockées dans des fichiers shapefile. Le schéma du cube est présenté à la figure 4-8 et vous trouverez en

Annexe A le schéma en étoile ainsi que les diagrammes de membres des dimensions.

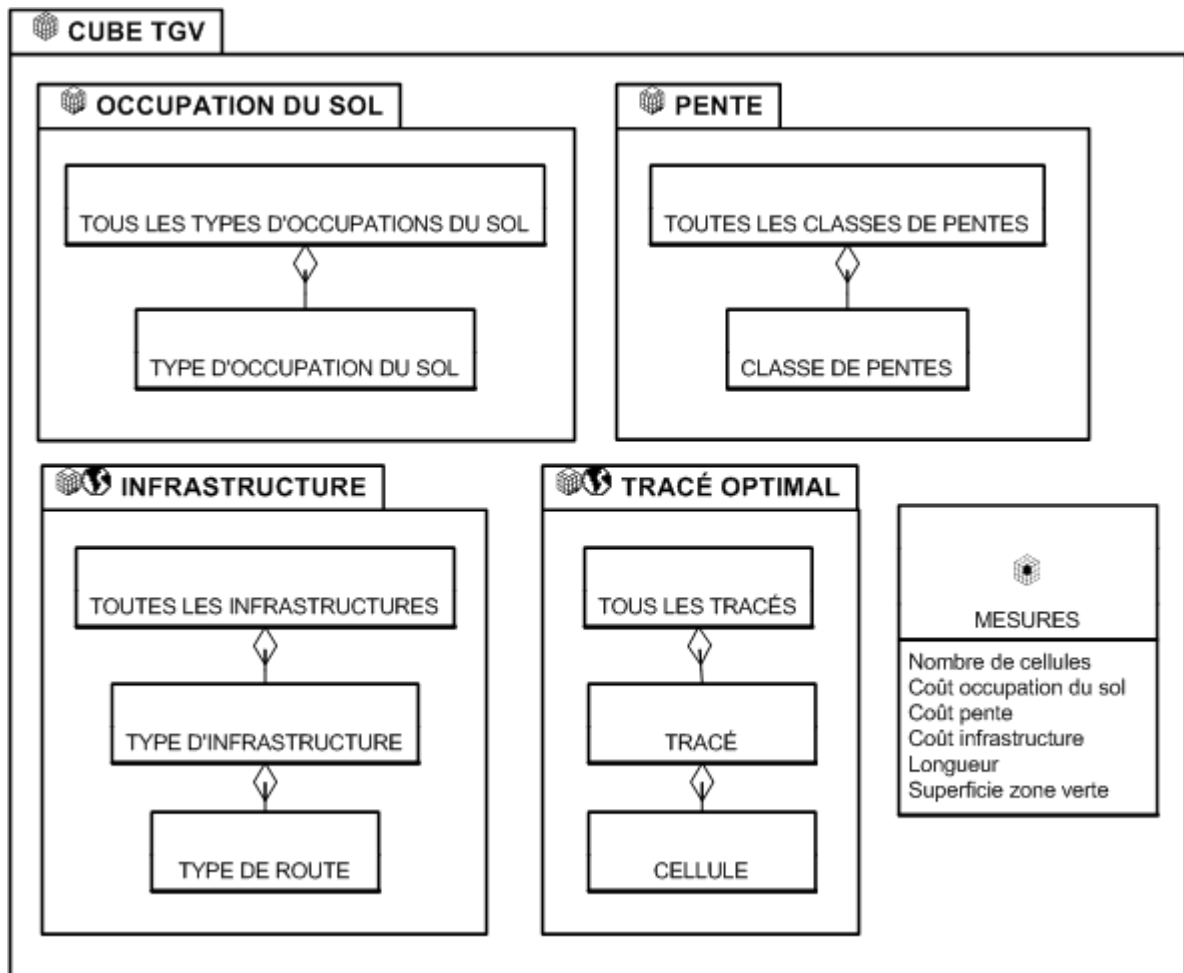


Figure 4-8 Schéma du cube TGV

Détail des différentes dimensions

La dimension thématique « Occupation du sol » contient une classification des principaux types d'occupations du sol (forêt, agriculture, hydrographie, urbain et autre). Le type de sol de chacune des cellules (faits détaillés) a été déterminé en effectuant une classification de l'image Landsat 7 à l'aide du logiciel PCI. Un rééchantillonnage de l'image classifiée a ensuite été réalisé à l'aide du logiciel ArcGIS d'ESRI version 9.2 pour obtenir des cellules de 60mx60m.

La dimension thématique « pente » est une classification de la pente en trois classes soit : <4%, 4% - 7% et >7%. La pente des cellules a été obtenue à l'aide d'une analyse de pente réalisée sur un MNT lui-même réalisé à l'aide de données d'élévation provenant de la BNDT.

La dimension spatiale non-géométrique « Infrastructure » permet d'identifier les cellules où il y a une infrastructure existante comme une route ou une ligne électrique. La dimension est une dimension spatiale non-géométrique puisque les membres ne possèdent pas de primitive géométrique. Les données proviennent de la BNDT et sont en format vectoriel. Elles ont ainsi été rasterisées pour être implantées dans le cube. Nous avons ensuite vectorisé la rasterisation dû aux limites de la technologie utilisée.

Détail des différentes mesures

Les **mesures de coûts** (*Coût occupation du sol, Coût pente et Coût infrastructure*) sont obtenues en effectuant des reclassifications des couches d'occupation du sol, de pente et d'infrastructure selon le coût attribué pour chaque type d'occupation du sol, pour chaque classe de pente et pour chaque infrastructure traversée. Le coût global pour la construction peut être calculé à la volée dans une mesure calculée en effectuant la somme des mesures de coûts.

La **mesure de longueur** correspond à la résolution des cellules soit 60m.

La **mesure superficie de zone verte** est le résultat d'une rasterisation des polygones de zonages agricole, suivi d'une reclassification attribuant 3600m² pour les cellules en zone verte et 0m² pour les cellules hors zones.

Le cube ainsi construit contient à la fois des phénomènes discrets (infrastructure) et continus (occupation du sol, pente) et permet une analyse par couverture de ces informations.

4.2.2.3 Limites, difficultés rencontrées et solutions apportées

Deux difficultés ont été rencontrées lors de l'implantation de ce cube. Premièrement, la dimension spatiale « tracé optimal » est non-stricte, puisqu'elle contient une relation multiple entre le niveau cellule et le niveau tracé. Les tracés sont composés de plusieurs cellules et puisqu'ils peuvent se croiser dans l'espace, une cellule peut appartenir à plus d'un tracé. Cette relation peut être matérialisée à l'aide d'une des solutions présentées à l'expérimentation 1 (cf. section 4.2.1.3). Ce qui nous intéresse plus particulièrement dans ce cube concerne les statistiques des tracés potentiels de TGV. La solution 2 aux difficultés rencontrées lors de l'expérimentation 1 (création d'une relation 1 : N) n'est donc pas appropriée et conseillée, car il y aura trop de pertes d'informations, si on associe les cellules appartenant à plusieurs tracés à un seul d'entre eux. Bien qu'il soit possible d'utiliser la solution 3 (dupliquer les cellules), l'application de cette solution dans un cube matriciel implique une plus grande redondance. Étant donné qu'un membre du niveau détaillé de la dimension spatiale « tracé optimal » correspond à une cellule et que cette cellule correspond également à un fait détaillé du cube, si on multiplie les cellules dans la dimension, on doit aussi multiplier les faits de la table de faits. Le volume de la table de faits est ainsi augmenté et l'agrégation des données selon les autres dimensions doit tenir compte des doublons dans les faits. Malgré cela, nous avons tout de même utilisé la solution 3 pour matérialiser la relation N-N dans la dimension spatiale « tracé optimal » étant donné que peu de cellules étaient concernées par cette relation. En fait, cela a nécessité la création de 54 nouveaux membres dans la dimension et le même nombre de faits dans la table de faits.

La deuxième difficulté rencontrée concerne également une relation multiple, mais cette fois-ci entre la table de faits et la dimension « infrastructure ». Les infrastructures telles que les routes et les lignes électriques peuvent s'intersecter. Les cellules correspondant aux faits détaillés peuvent ainsi être associées à plus d'un type d'infrastructure de la dimension « infrastructure ». Tout comme les dimensions non-strictes, les relations multiples entre faits et

dimension nécessitent des règles d'implantation particulières. Il existe plusieurs stratégies selon que la relation N-N est finie (ex : 1-3) ou non (N-N). Le terme stratégie est simplement utilisé ici pour ne pas confondre avec les solutions d'implantations des hiérarchies non-strictes. Nous présentons ici les différentes stratégies telles que décrites par Bédard [Bédard et *al.*, 2006a].

Stratégie 1 : Créer dans la dimension l'ensemble des combinaisons possibles (ex. routière, hydro, routière-hydro).

Stratégie 2 : Multiplier les membres dans la dimension (ex. principal, secondaire). Cette stratégie peut toutefois introduire une dimension non-strictes.

Stratégie 3 : Multiplier les dimensions dans le cube (ex. une dimension « infrastructure routière », une dimension « infrastructure hydrographique »).

Stratégie 4 : Prioriser un membre de la dimension sur un autre (selon par exemple la superficie ou la longueur couverte par l'infrastructure).

Stratégie 5 : Exploiter la redondance d'information dans les faits (c.-à-d. ajouter un nouveau fait pour chaque infrastructure). Utiliser les faits multiples en considérant les doublons dans l'agrégation des données. Attention, dans un cube matriciel, il faut être prudent avec cette stratégie. Si ce cube contient plus d'une dimension spatiale géométrique vectorielle ou hybride, la gestion des doublons peut devenir très complexe dans ces cas.

Stratégie 6 : Implanter la relation en 1-N et aviser l'utilisateur par une mise en garde. Cette stratégie peut être utilisée si le nombre de fois où la relation N-N exploitée est faible et si le nombre calculé résultant varie peu par rapport au nombre réel.

Stratégie 7 : Utiliser une table-pont pour décrire directement la relation N-N entre les faits et les membres d'une dimension. Le serveur utilisé doit supporter ce type de table ce qui n'est pas le cas pour JMap Spatial OLAP.

Pour notre expérimentation, puisque la relation N-N entre la table de faits et la dimension « infrastructure » est finie (1-2), nous avons choisi d'utiliser la stratégie 1 en créant un nouveau membre dans la dimension qui correspond à la combinaison entre route et ligne électrique.

4.2.2.4 Exemples d'analyses

Le cube permet d'analyser le territoire couvert en fonction des critères imposés pour la construction d'un tracé de TGV. L'agrégation des cellules au niveau tracé potentiel de la dimension spatiale permet d'analyser les tracés potentiels de TGV et de les comparer selon leur longueur et leur coût (Figure 4-9). Il est également possible de déterminer la longueur des tracés par type d'occupation du sol ou par classe de pente. Ou encore de segmenter les tracés et de visualiser les différents segments affectés par des critères particuliers comme illustré sur la figure 4-10 où on représente les segments du tracé Sainte-Foy 1 en zone agricole où la pente est < 4%.

	LONGUEUR	Coût global
Lorette	10860	8504784
Sainte-Foy1	15480	20312424
Sainte-Foy2	14940	16571628

Figure 4-9 Longueur et coût global des tracés potentiels de TGV

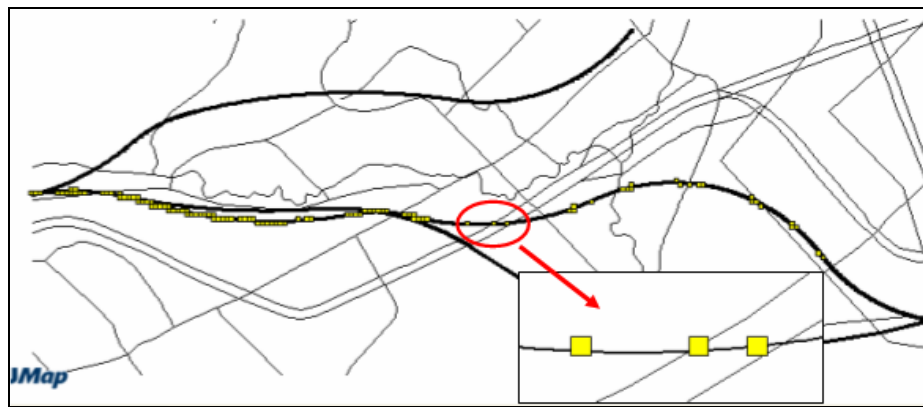


Figure 4-10 Segments du tracé Sainte-Foy 1 en zone agricole où la pente est < 4%

4.2.3 Expérimentation exploratoire 3

Les deux premières expérimentations ont permis de tester les concepts de dimensions spatiales matricielles et de cube matriciel. Cette troisième expérimentation explore maintenant le **potentiel d'analyse spatiale matricielle à la volée**. Ainsi, nous avons testé l'implantation d'un opérateur de Map Algebra dans le serveur OLAP Open Source Mondrian (<http://mondrian.pentaho.org/>) permettant d'effectuer une analyse spatiale à la volée à même un cube de données. Le serveur Mondrian est complètement écrit en Java et permet d'interroger des cubes de données à l'aide du langage MDX (Multidimensional Expressions) [Spofford et *al.*, 2006]. Ce langage est une variante multidimensionnelle du langage SQL utilisé pour interroger des bases de données transactionnelles.

Les données et le cube

Les données utilisées dans le cadre de cette expérimentation sont celles des inventaires forestiers de la forêt Montmorency. Ces données sont composées de peuplements forestiers et de leurs caractéristiques telles que l'essence, la densité, l'âge, la hauteur, la pente et les perturbations. Le cube de données qui a été construit est un **cube matriciel** très simple dans lequel on retrouve 2 dimensions, une mesure et une mesure calculée (Figure 4-11). Les données du cube sont stockées dans une base de données PostgreSQL 8.2.

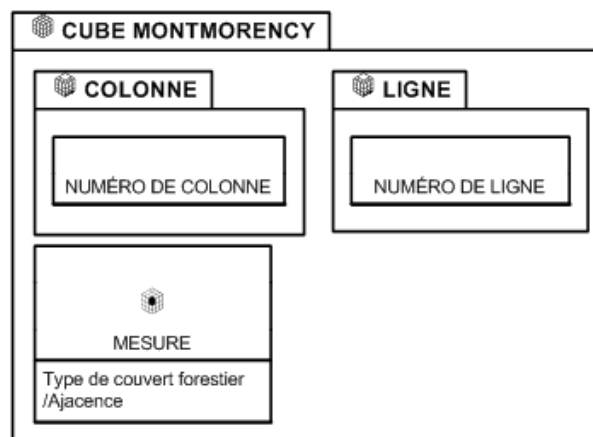


Figure 4-11 Schéma du cube Montmorency

Description des mesures et des dimensions

Les faits détaillés du cube sont des cellules d'une grille matricielle et les dimensions « Colonne » et « Ligne » représentent respectivement les numéros de ligne et de colonne de ces cellules dans la grille. La mesure *Type de couvert forestier* correspond au couvert forestier de chacune des cellules soit : résineux, mélangé, feuillu, hydrographie ou inconnu. Ces informations ont été obtenues en effectuant une rasterisation des peuplements forestiers selon leur essence. Finalement, la mesure calculée *Adjacence* correspond au résultat d'une opération de Map Algebra réalisée à partir de la mesure *type de couvert forestier*. Une mesure calculée est obtenue en effectuant une opération sur une ou plusieurs mesures existantes [Spofford et al., 2006]. L'opérateur de Map Algebra utilisé est l'opérateur nommé « Focal Rating » [Tomlin, 1990]. Cet opérateur est de type Focal c'est-à-dire qu'il tient compte de la valeur de la cellule courante et de celles de ses voisines (cf. section 2.3.3). Il permet d'effectuer une analyse d'adjacence, qui se résume à identifier s'il y a ou non adjacence entre 2 types de couvert forestier.

Étant donné que le serveur Mondrian est un serveur OLAP Open Source, il est possible de supprimer, de modifier ou d'ajouter des fonctionnalités pour répondre à nos propres besoins. Nous avons donc ajouté une nouvelle fonctionnalité permettant la réalisation de l'opérateur « Focal Rating ». Cet opérateur a été implanté dans le serveur comme une fonction de type « UserDefinedFunction » et nécessite en entrée la valeur de la cellule courante et les valeurs de ses 8 cellules voisines. Le code de cette fonction se retrouve en Annexe C. L'appel de la fonction s'effectue à l'aide d'une requête MDX également présentée en Annexe C. L'appel d'une telle fonction parcourt le cube fait par fait et retourne, pour chacun de ces faits, une valeur pour la mesure calculée. Puisque le cube est un cube matriciel et que les faits sont des cellules d'une grille matricielle, on peut facilement comparer la réalisation d'une telle fonction dans un cube à la réalisation d'une opération de Map Algebra dans un modèle matriciel. En effet, dans un modèle matriciel, une opération de Map

Algebra s'effectue cellule par cellule et retourne une nouvelle couche matricielle correspondant à une nouvelle valeur pour chacune des cellules. La réalisation d'une telle fonction dans un cube matriciel s'effectue ainsi exactement de la même manière que dans un modèle matriciel, la couche résultante étant simplement implantée en une mesure calculée.

Les dimensions « Ligne » et « Colonne » du cube permettent d'aller chercher les valeurs d'entrées nécessaires à la réalisation de la fonction en identifiant la cellule courante et en allant chercher les valeurs de ses 8 voisines en additionnant 1 ou en soustrayant 1 aux numéros de ligne et de colonne de cette cellule. L'opérateur Focal Rating ainsi implanté permet de déterminer les endroits où il y a adjacence entre 2 types de couvert forestier identifiés par l'utilisateur. Il parcourt le cube en prenant en entrée la valeur de la mesure *Type de couvert forestier* pour la cellule courante ainsi que pour ses 8 voisines et retourne 1 s'il y a adjacence et 0 s'il n'y a pas adjacence. Il est ainsi possible d'identifier par exemple les endroits où il y a adjacence entre le type de couvert forestier résineux et le type hydrographie. Étant donné que le serveur Mondrian ne gère pas la composante spatiale des données, le cube créé ne contient pas de géométrie et les résultats ne peuvent pas être visualisés cartographiquement dans le cube. Toutefois, il est possible d'exporter la mesure résultante et d'utiliser un visualisateur ou un SIG pour représenter le résultat (ex. logiciel ArcGIS d'ESRI).

Cette expérimentation a permis de tester la réalisation d'une analyse spatiale matricielle à la volée dans un cube de données. La réalisation de l'opérateur Focal Rating sur un cube contenant 21 917 faits s'effectue en 4 secondes (à l'aide d'un Intel T2500 avec 1GB de mémoire vive) ce qui est conforme avec les exigences des outils décisionnels [Marchand, 2004]. Toutefois, les algorithmes utilisés devront être optimisés afin d'améliorer les performances sur de plus gros volumes de données. Des tests ont été réalisés sur des cubes contenant 87 899 et 351 596 faits et les temps de traitements ont été respectivement de 17 secondes et 2 minutes. Dans le cadre de cette expérimentation, la réalisation d'une analyse spatiale matricielle à la volée

démontre bien que la structure matricielle apporte un réel potentiel pour améliorer les analyses spatiales interactives dans les outils SOLAP.

4.3 Expérimentation formelle

Les concepts théoriques ayant été testés dans le cadre des expérimentations exploratoires, l'expérimentation formelle permet maintenant de valider leurs utilités à répondre à un besoin réel et actuel. Cette expérimentation cadre dans un projet de recherche entre la Chaire de recherche industrielle en bases de données géospatiales décisionnelles et la Recherche et Développement pour la Défense Canada - Valcartier (RDDC-Valcartier). L'expérimentation s'inscrit dans la 3^e phase de ce projet, faisant ainsi suite à l'analyse théorique et à la définition des spécifications d'une infrastructure de découverte et d'accès aux données géospatiales (phase 1- livrable 3-2007), ainsi qu'à la mise en place d'une gestion multidimensionnelle des métadonnées (phase 2- livrable 4-2007). La 3^e phase sert, quant à elle, de base conceptuelle au développement de la seconde version du prototype d'outil décisionnel GÉOLAP prévu pour 2008. Dans le cadre de cette expérimentation, nous avons été impliqués dans la modélisation des prototypes qui ont servi à démontrer l'utilité de la structure matricielle pour le développement de la 2^e version de l'application GÉOLAP.

4.3.1 Problématique actuelle

Le projet GÉOLAP a pour objectif la gestion et l'analyse d'un inventaire cartographique et son exploitation par ses utilisateurs. La première version de l'application GÉOLAP permet d'exploiter l'information sur les métadonnées des documents cartographiques de l'inventaire dans un environnement BI (Business Intelligence) exploitant la technologie OLAP et intégrant des fonctionnalités SIG. Elle permet également l'analyse de documents spatialement localisés en fonction de leurs métadonnées telles que le type de produit, l'échelle cartographique, la résolution, l'année de validité et l'année de production. En novembre 2007, 88 156 documents de l'inventaire étaient

stockés dans l'application. D'autres saisies sont à venir pour compléter cet inventaire, ce qui augmentera considérablement le nombre de documents pour la 2^e version de l'application GÉOLAP. L'envergure et la diversité d'un tel inventaire de documents engendre une problématique pour la visualisation de la couverture des documents sur le territoire. Les collections cartographiques peuvent posséder des découpages réguliers continus ou irréguliers. De plus, plusieurs collections peuvent porter sur la même région ce qui fait en sorte que les découpages cartographiques se superposent sur le territoire. Des techniques de couleur et de transparence sont actuellement utilisées dans l'outil GÉOLAP pour permettre d'illustrer la couverture cartographique des documents sur le territoire. Cependant, l'appréciation du territoire couvert par les documents est difficilement évaluable par l'utilisateur étant donné le nombre de documents se chevauchant et la perte de perception due à l'enchevêtrement des lignes et des couleurs. Il devient donc nécessaire d'utiliser de nouvelles techniques ou de nouvelles approches pour améliorer la perception visuelle des collections de documents de l'application GÉOLAP. La figure 4-12 illustre un exemple de la couverture spatiale de plusieurs documents tels qu'ils sont actuellement présentés dans l'application GÉOLAP.

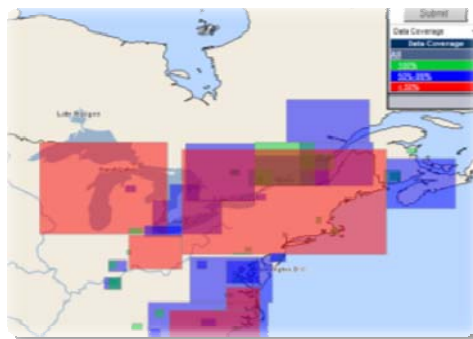


Figure 4-12 Étendue cartographique des documents tels que présentés dans GÉOLAP (Source : [Proulx et *al.*, 2008])

4.3.2 Approches proposées

Dans le cadre de ce projet, différentes approches ont été proposées pour améliorer la perception visuelle de l'information dans l'application GÉOLAP. L'approche qui nous intéresse propose d'indexer l'information selon une

structure matricielle. En éliminant la superposition visuelle, l'utilisation du matriciel permettra de mieux représenter l'information de manière cartographique. Les concepts élaborés au chapitre 3 ont ainsi été utilisés pour l'intégration d'une structure de données matricielle dans des prototypes SOLAP basés sur les données de l'application GÉOLAP. Les prototypes développés dans le cadre de cette expérimentation ont ainsi été réalisés à l'aide de l'outil JMap Spatial OLAP. Les résultats de cette expérimentation serviront pour l'implantation de la deuxième version de l'application GÉOLAP prévue pour 2008.

Le type d'analyse réalisé à l'aide de l'application est une analyse par objets, c'est-à-dire qu'on analyse des documents qui sont spatialement localisés en fonction de leurs métadonnées telles que le type de produit, l'échelle cartographique, la résolution, l'année de validité et l'année de production. Dans un contexte d'analyse par objets, le type de cube à utiliser est un cube d'objets. L'intégration du matriciel dans un cube d'objets s'effectue uniquement par l'utilisation de dimensions spatiales. L'équipe de la chaire de recherche industrielle en bases de données géospatiales décisionnelles a ainsi proposé différentes solutions pour la gestion de la dimension spatiale du territoire.

Deux prototypes de cube d'objets ont été construits à l'aide de JMap Spatial OLAP, l'un incluant une dimension spatiale géométrique matricielle et l'autre une dimension spatiale géométrique hybride. Les 2 dimensions pourraient éventuellement être implantées dans un même cube. Ces dimensions permettent d'indexer les documents sur un découpage régulier composé de cellules de base. Il est ainsi possible de rapidement évaluer le nombre de documents couvrant le territoire demandé par la sélection d'un groupe de cellules de base et d'ensuite agréger l'information selon un découpage régulier de résolution plus grossière (dimension géométrique matricielle) ou selon un découpage prédéfini tel que celui des pays et des continents (dimension géométrique hybride). Dans un contexte décisionnel, la représentation exacte des objets n'est souvent pas aussi importante que dans un contexte

transactionnel. Dans ce contexte, l'indexation des documents sur des cellules régulières ne constitue donc pas une limite pour l'analyse. La **dimension géométrique matricielle** est composée de trois niveaux de cellules régulières, soient un niveau composé de 1 800 cellules de 6 degrés, un second niveau composé de 16 200 cellules de 2 degrés et un niveau fin composé de 64 800 cellules de 1 degré. La **dimension géométrique hybride** quant à elle est composée de quatre niveaux soient un niveau monde (vectoriel), un niveau continent (vectoriel) avec le Canada et les États-Unis représentés individuellement, un niveau pays (vectoriel) avec les provinces canadiennes et les États américains utilisés pour l'Amérique du Nord et un niveau fin (matriciel) composé des cellules de 1 degré (voir schémas en Annexe A). La résolution des cellules de base a été fixée à 1 degré suite à une analyse des relations possible entre les documents et les cellules de bases (Tableau 4-1).





Possibilités	Illustration	Relation	Indexation
Une cellule contient entièrement un seul document		1 : 1	Indexation d'un document sur une cellule
Une cellule contient entièrement plus d'un document		1 : N	Indexation de plusieurs documents sur une cellule
Une cellule contient en partie un seul document		N : 1	Indexation d'un document sur plusieurs cellules
Une cellule contient en partie plusieurs documents		N : N	Indexation de plusieurs documents sur plusieurs cellules

Tableau 4-1 Relations possibles entre les cellules et les documents (Source : [Proulx et *al.*, 2008])

Le fichier de données géométriques contient les couvertures de plus de 80 000 documents de formes irrégulières se superposant sur le territoire. L'indexation de ces documents sur des cellules régulières implique inévitablement des relations N-N (relations multiples entre la table de faits et les dimensions spatiales). La matérialisation de ces relations a été réalisée à l'aide de la stratégie 5 présentée à la section 4.2.2.3 c'est-à-dire que les faits détaillés ont été multipliés dans la table de faits. Donc, plus il existe de relations multiples

entre les éléments, plus la table de faits contenant les 80 000 documents est multipliée. La résolution des cellules de base a ainsi un impact important sur la taille de la table de faits. Le niveau cellule de base est utilisé dans les 2 prototypes, puisqu'il forme le niveau fin tant de la dimension géométrique matricielle que de la dimension géométrique hybride. La matérialisation de la relation N-N a donc été traitée de la même manière dans les 2 cas. Pour des cellules de 1 degré, l'ensemble des faits détaillés des cubes se chiffre à 800 000 enregistrements représentant les combinaisons possibles entre les 80 000 documents et les 64 800 cellules de 1 degré. Pour le calcul des agrégations, il a été nécessaire de tenir compte de ces doublons lors des calculs aux niveaux supérieurs des dimensions.

La dimension géométrique hybride contient en plus une relation N-N entre le niveau cellule et le niveau pays. Un pays peut être composé de plus d'une cellule et une cellule peut chevaucher plus d'un pays. Cette relation a ici été matérialisée à l'aide de la solution 3 présentée à la section 4.2.1.3, ce qui a nécessité la multiplication des cellules appartenant à plus d'un pays dans la dimension et dans le fichier shapefile contenant les géométries des cellules. Les schémas des 2 cubes sont présentés en Annexe A et nous présentons ici quelques exemples d'analyses réalisables grâce à l'intégration du matriciel dans les dimensions des cubes.

Cube d'objets avec dimension géométrique matricielle

Ce cube permet de visualiser l'ensemble du territoire en cellules régulières de différentes grosseurs et d'apprécier le nombre de documents couvrant ces cellules selon différents critères (type de collection, date de validité et date d'édition). La figure 4-13 illustre le nombre de documents sur le découpage mondial représenté par le niveau cellule de 6 degrés. La figure suivante (figure 4-14) illustre l'Amérique du Nord représentée selon les 3 niveaux de cellules (6 degrés, 2 degrés et 1 degré).

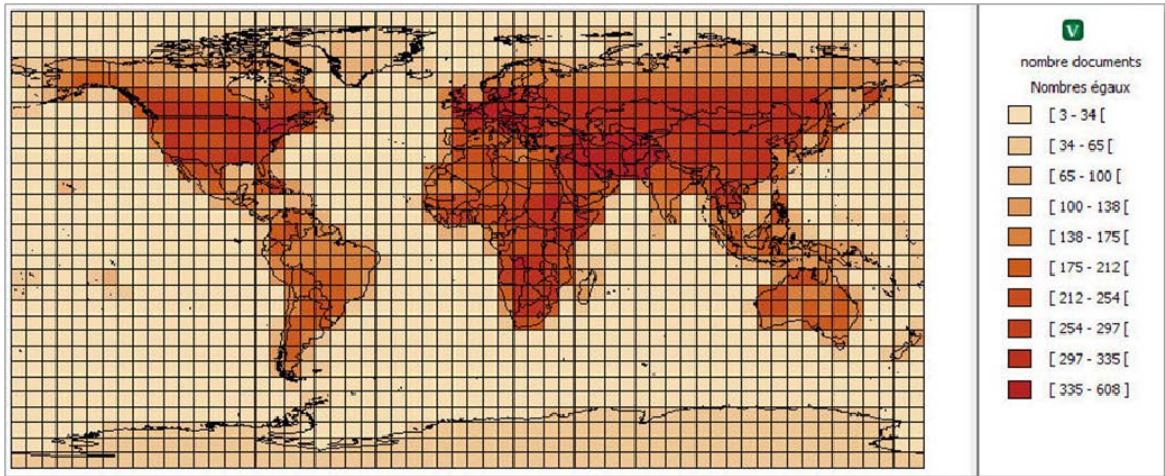


Figure 4-13 Région mondiale représentée par les cellules de 6 degrés (Source : [Proulx et *al.*, 2008])

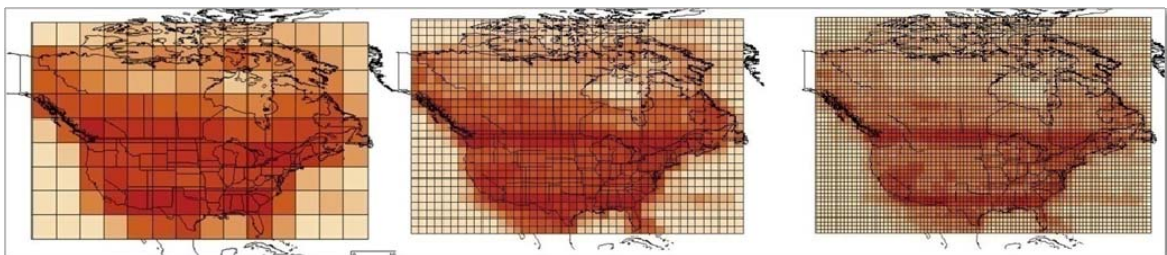


Figure 4-14 L'Amérique du Nord représentée par les cellules de 6 degrés, 2 degrés et 1 degré (Source : [Proulx et *al.*, 2008])

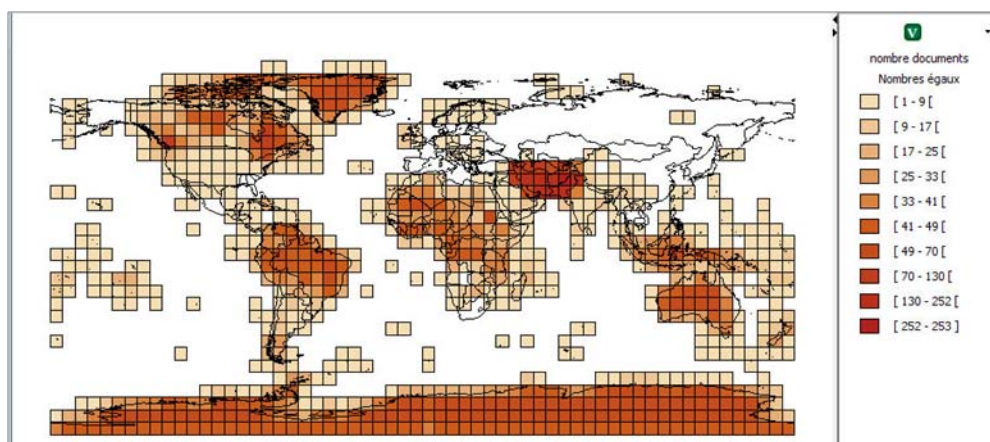


Figure 4-15 Filtre sur la couverture mondiale pour les documents de date de validité > 2005 (Source : [Proulx et *al.*, 2008])

Il est également possible de filtrer les documents selon leurs caractéristiques à l'aide des autres dimensions du cube tel que présenté à la figure 4-15. Un filtre est ici appliqué sur la couverture mondiale pour les documents de date de validité > 2005

L'utilisation d'un découpage par cellules permet de représenter la distribution des documents sur le territoire beaucoup plus facilement que le permet l'application GÉOLAP actuelle. Dans l'application actuelle, la superposition des documents limite la perception de la quantité réelle représentée sur le territoire. L'affichage du nombre de documents sur une couche matricielle tel que proposé est beaucoup plus efficace. Par contre, l'étendue de chaque document n'est plus disponible, ce qui ne pose pas de problème dans ce cas, car l'essentiel est de connaître la couverture mondiale des documents afin de voir les zones moins bien couvertes, et non le détail géométrique de chacun des documents. Il serait cependant possible d'insérer une nouvelle fonctionnalité permettant, à partir d'une cellule, de sélectionner les documents inclus dans cette cellule et de voir leur étendue. L'utilisation d'une dimension géométrique matricielle permet également de faire des regroupements ad hoc c'est-à-dire sans restrictions de découpages géographiques prédéfinis tels que celui des pays. Toutefois, la sélection des cellules d'une zone particulière est difficile, par exemple les cellules composant le Canada. Dans l'arbre de la dimension, les cellules ne sont pas regroupées selon un regroupement nominal ou géographique, elles apparaissent simplement par numéro de cellules. Cet inconvénient est corrigé par l'utilisation d'une dimension géométrique hybride telle qu'implantée dans le deuxième cube.

Cube d'objets avec dimension géométrique hybride

Ce cube permet de visualiser le territoire selon un découpage géographique vectoriel et selon un découpage régulier formé de cellules de 1 degré. La figure 4-16 représente le nombre de documents agrégés au niveau pays et la figure 4-17 présente le nombre de documents au niveau cellule pour un ensemble de pays sélectionnés.

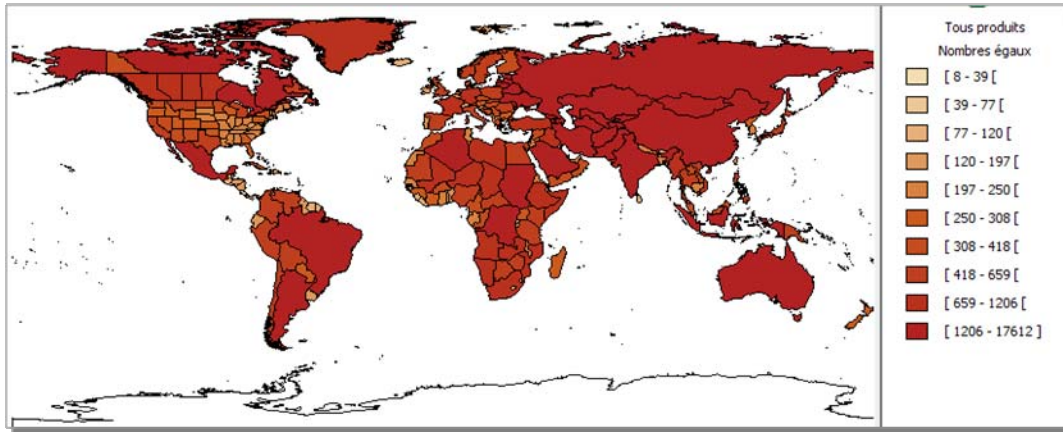


Figure 4-16 Nombre de documents selon le découpage des pays (Source : [Proulx et *al.*, 2008])

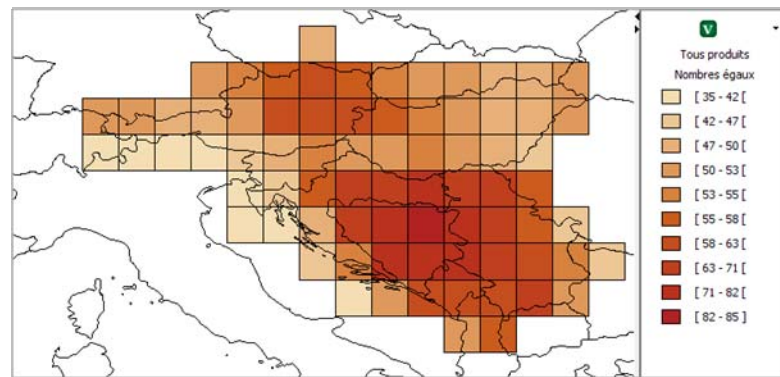


Figure 4-17 Nombre de documents pour un ensemble de pays (Source : [Proulx et *al.*, 2008])

La structuration de la dimension géométrique hybride facilite la sélection des cellules par l'identification de régions dans la dimension et permet d'apprécier plus facilement la quantité de documents disponibles sur ces régions.

Les prototypes de cubes d'objets incluant des dimensions spatiales géométriques matricielle et hybride construits dans le cadre de cette expérimentation ont permis de démontrer l'utilité d'éventuellement implanter une structure de données matricielle dans les dimensions spatiales de la prochaine version de l'application GÉOLAP. L'implantation du matriciel nécessite quelques traitements supplémentaires permettant l'indexation des

documents sur les cellules et il augmente considérablement le volume de la table de faits. Par contre, il permet de représenter la distribution des documents sur le territoire beaucoup plus facilement que le permet l'application GÉOLAP actuel. Il permet également de faire des regroupements ad hoc sans restriction de découpages géographiques prédéfinis. La relation existante entre les documents et les cellules permettrait éventuellement d'introduire d'autres données portant sur des phénomènes sociaux économiques tels que la population ou les zones à risque et permettrait ainsi d'analyser l'information de l'inventaire en fonction d'autres phénomènes pour potentiellement découvrir des connaissances géospatiales d'intérêts. Dans le cadre de ce projet, l'analyse du potentiel d'un cube matriciel a également été réalisée, mais puisque l'utilisation d'une telle approche est très différente de l'exploitation actuelle de l'inventaire cartographique et qu'elle n'apportait rien de plus à l'application, elle n'a pas été retenue.

Les informations et les exemples qui ont servi à la description de cette expérimentation ont été tirés du rapport « Amélioration de la représentation visuelle et de la gestion des découpages cartographiques dans l'application GÉOLAP » [Proulx et *al.*, 2008]. Nous avons, dans cette section, fait un résumé de ce rapport, pour plus de détails sur la construction des cubes, le lecteur est invité à consulter le rapport complet.

4.4 Conclusion

Dans ce chapitre, nous avons testé et validé les principaux concepts élaborés au chapitre 3 permettant la gestion et l'intégration du matriciel dans des cubes de données avec les technologies existantes. Trois expérimentations exploratoires ont permis de démontrer la faisabilité d'exploiter la structure matricielle dans un cube de données et ont également aidé à développer nos concepts théoriques. Dans le cadre de ces expérimentations, nous avons testé les deux approches permettant l'intégration du matriciel dans un cube de données soit dans les dimensions spatiales et dans les faits. Nous avons

premièrement testé l'intégration du matriciel dans les dimensions spatiales en construisant un cube d'objets composé d'une dimension géométrique matricielle et d'une dimension géométrique hybride (expérimentation exploratoire 1). L'intégration du matriciel dans les faits a par la suite été testée. Nous avons construit un cube matriciel composé d'une dimension spatiale géométrique hybride (expérimentation exploratoire 2) et un deuxième cube matriciel cette fois-ci sans dimension spatiale qui a permis d'implanter un opérateur d'analyse spatiale matricielle (expérimentation exploratoire 3). L'ensemble des combinaisons possibles entre les types de dimensions spatiales et les types de cubes n'ont pas été testés ici. Par contre, les trois expérimentations exploratoires ont été suffisantes pour démontrer qu'il est possible d'intégrer le matriciel dans les faits d'un cube, qu'il est possible d'intégrer le matriciel dans les dimensions spatiales d'un cube peu importe son type (d'objets ou matriciel), qu'il est possible de combiner le matriciel et le vectoriel dans les dimensions spatiales (dimension géométrique hybride), qu'il est possible d'utiliser le vectoriel dans les dimensions d'un cube matriciel, et finalement qu'il est possible de réaliser des analyses spatiales à la volée dans un cube matriciel. Une expérimentation formelle a par la suite permis de valider le potentiel d'intégrer le matriciel dans un cube de données en démontrant son potentiel à répondre à un besoin réel et actuel. Pour cette expérimentation, le matriciel a été implanté dans les dimensions spatiales de deux prototypes SOLAP développés à partir des données de l'application GÉOLAP. Ces prototypes ont permis de démontrer le potentiel du matriciel pour améliorer la représentation visuelle des données de la prochaine version de l'application GÉOLAP.

La réalisation de ces expérimentations à l'aide de l'outil JMap Spatial OLAP a également fait ressortir quelques limites technologiques liées à l'intégration du matriciel dans les dimensions spatiales. L'intégration du matriciel dans les dimensions spatiales implique la création de plusieurs membres puisque chaque cellule d'une grille matricielle correspond à un membre de la dimension. Toutefois, l'outil utilisé actuellement est limité quant au nombre d'enfants des membres des dimensions. Cette limite influence les requêtes sur

la table de faits et l'arbre des membres. Ce nombre maximal n'est actuellement pas connu, toutefois nous savons qu'il est influencé par le type d'identifiant des membres. Cette limite est surtout problématique pour les dimensions spatiales géométriques hybrides dont les membres vectoriels possèdent de grandes superficies se décomposant à un niveau inférieur en cellules matricielles de résolution fine. Pour ces cas, il est nécessaire de définir des niveaux intermédiaires dans les dimensions de façon à ce que le nombre de membres par branches soit inférieur à cette limite. Une deuxième limite purement technologique réside dans le rafraîchissement de l'arbre des dimensions de l'outil SOLAP. En fait, le rafraîchissement de l'arbre des dimensions peut s'avérer assez long lorsqu'une requête impliquant plusieurs membres localisés sous différentes branches est réalisée. C'est notamment le cas pour les requêtes impliquant l'ouverture de l'arbre d'une dimension géométrique matricielle. Ces limites technologiques sont ainsi à prendre en considération lors de la réalisation d'applications à l'aide de l'outil JMap Spatial OLAP.

Les concepts ayant été testés et validés, le prochain chapitre initie l'étude du potentiel matriciel pour l'analyse spatiale dans les outils SOLAP.

CHAPITRE 5 POTENTIEL DU MATRICIEL POUR L'ANALYSE SPATIALE DANS LES OUTILS SOLAP

5.1 Introduction

Les concepts fondamentaux permettant de concevoir des cubes supportant des données matricielles ont été définis, testés et validés dans les chapitres précédents. Basé sur ces concepts, il est maintenant possible d'initier l'étude du potentiel de la structure matricielle à supporter l'analyse spatiale dans les outils SOLAP. Nous identifions ici des possibilités d'exploitation d'analyses spatiales matricielles dans les outils SOLAP en fonction du type de cube utilisé, nous démontrons à l'aide d'exemples la réalisation de certaines de ces analyses dans un SOLAP et finalement, nous comparons ces possibilités avec celles actuellement réalisables à l'aide d'une structure vectorielle. Cette étude n'est pas exhaustive, c'est-à-dire qu'elle repose uniquement sur les expérimentations que nous avons réalisées au chapitre précédent, toutefois elles sont suffisantes pour avancer quelques résultats et pour démontrer le potentiel réel du matriciel à enrichir les analyses spatiales dans les outils SOLAP.

5.2 Possibilités d'exploitation des analyses spatiales matricielles dans un SOLAP

La situation des outils SOLAP en termes d'analyses spatiales est actuellement limitée. Nous avons vu au chapitre 1 que cette limite est principalement due à la structure de données vectorielle présentement utilisée dans ces outils. En présence de grandes quantités de données ou d'analyses agrégatives typiques des applications SOLAP, l'analyse spatiale vectorielle nécessite des opérateurs

complexes et lourds à exécuter. Cette complexité restreint le potentiel d'analyses interactives dans les outils SOLAP. Les analyses doivent ainsi être précalculées pour être exploitées. La figure 5-1 illustre les possibilités d'exploitation d'analyses spatiales vectorielles dans un SOLAP.

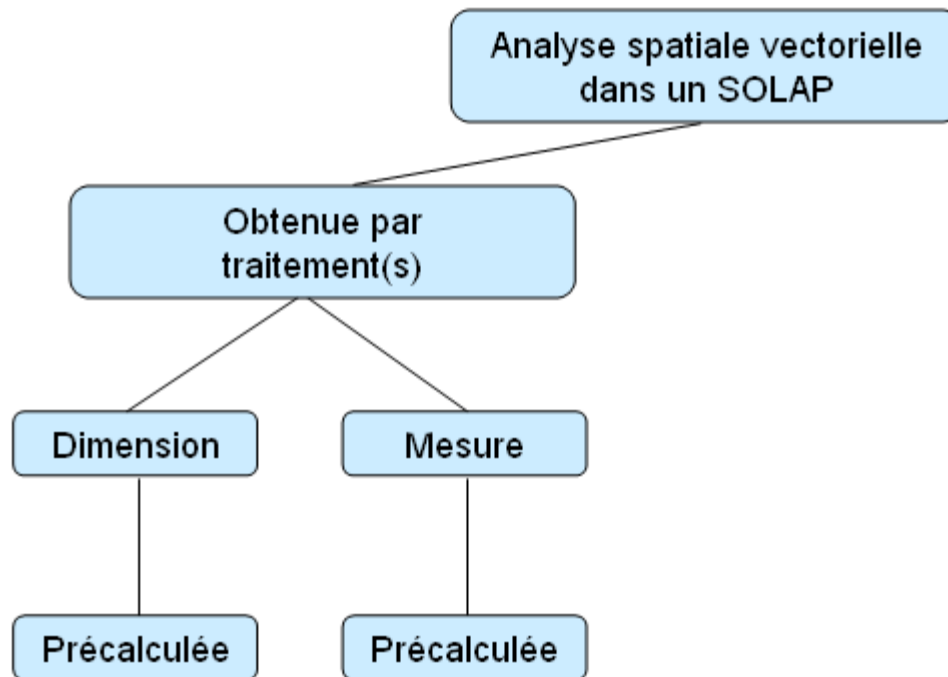


Figure 5-1 Possibilités d'exploitation des analyses spatiales vectorielles dans un SOLAP

Les résultats d'analyses spatiales vectorielles peuvent être exploités dans un cube soit en dimension, comme dans les travaux de Marchand [Marchand, 2004] où les analyses vectorielles sont exploitées dans une dimension d'opérateurs topologiques spatio-temporels (cf. section 2.4.2), soit en mesure spatiale. Les analyses vectorielles sont toujours précalculées et elles nécessitent souvent de nombreux traitements [Marchand, 2004].

Puisque les concepts théoriques permettant l'intégration de la structure matricielle dans les cubes spatiaux ont été définis, testés et validés, il est maintenant possible d'initier le potentiel du matriciel pour l'analyse spatiale dans les outils SOLAP. Les expérimentations du chapitre 4 ont d'ailleurs permis

d'identifier les possibilités d'exploitation d'analyses spatiales matricielles dans un SOLAP. La figure 5-2 illustre ces possibilités.

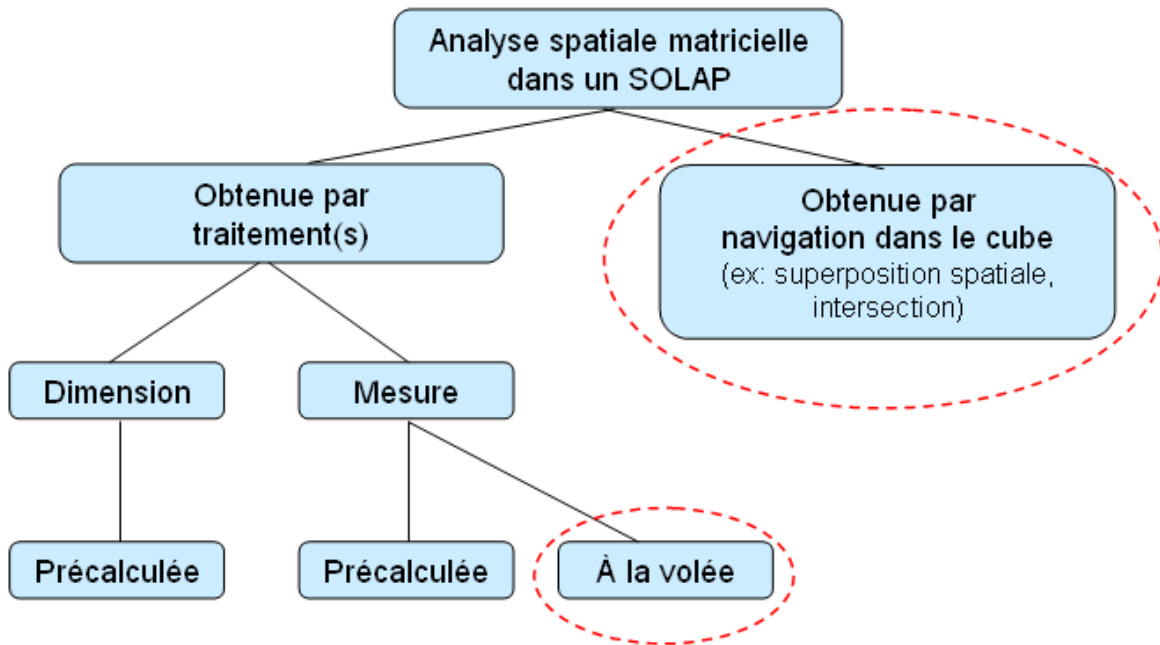


Figure 5-2 Possibilités d'exploitation des analyses spatiales matricielles dans un SOLAP

Comparativement aux possibilités d'exploitation des analyses vectorielles où les analyses sont obtenues uniquement par traitements, le matriciel permet également d'exploiter certaines analyses, comme des superpositions spatiales ou des intersections spatiales, par simple navigation dans le cube. Puisque les opérateurs d'analyses matricielles sont simples, il est en plus possible de réaliser des analyses à la volée (par opposition à précalculées) dans un cube. Ainsi, dans un cube, les résultats d'analyses spatiales matricielles peuvent être obtenus soit par traitements, soit par navigation dans le cube. Les analyses obtenues par traitements sont des analyses qui ont nécessité la réalisation d'opérations sur des données matricielles, tandis que les analyses obtenues par navigation sont obtenues simplement par le croisement de certaines dimensions dans un cube. Nous expliquons dans la suite de cette section ces

différentes possibilités, tout en nous référant à des exemples d'analyses réalisées lors des expérimentations décrites au chapitre 4.

Les sections 5.2.1 à 5.2.3 présentent tout d'abord les **analyses obtenues par traitements**. Ces analyses sont exploitables soit dans les **dimensions** (section 5.2.1), soit dans les **mesures** d'un cube (section 5.2.2 et 5.2.3). La section 5.2.4 présente ensuite les **analyses obtenues par navigation** dans un cube.

5.2.1 Les analyses obtenues par traitements au niveau des dimensions

L'exploitation d'une analyse matricielle **en dimension** nécessite le précalcul de celle-ci, c'est-à-dire que l'analyse doit être réalisée à l'extérieur de l'outil SOLAP. Les résultats sont ensuite implantés dans une dimension thématique ou une dimension spatiale descriptive. Les dimensions d'un cube sont composées de membres hiérarchisés permettant l'agrégation des mesures du cube. Lors de la conception d'un cube, ces membres sont préalablement identifiés et hiérarchisés. Les résultats d'analyses spatiales matricielles exploités dans une dimension devront ainsi être précalculés et classifiés pour former les membres de la dimension. Les résultats d'analyses matricielles sont de nouvelles couches matricielles pour lesquelles une nouvelle valeur est attribuée à chaque cellule de la grille. Puisque la classification de ces valeurs est une opération courante, il est donc simple de classier le résultat d'une analyse matricielle pour former des membres d'une dimension.

L'exploitation d'analyses matricielles en dimension est possible dans un cube matriciel ou dans un cube d'objets qui intègre le matriciel dans ses dimensions spatiales. Les analyses de distances de l'expérimentation 1 (cf. section 4.2.2.1) et l'analyse de pente de l'expérimentation 2 (cf. section 4.2.2.2) sont des exemples d'analyses matricielles exploitées dans les dimensions d'un cube (voir Annexe A pour les diagrammes de membres de ces dimensions). Elles permettent d'effectuer une sélection des éléments du cube en fonction d'un critère d'analyse spatiale tel que la distance à une école. Il est ainsi possible

d'effectuer des analyses telles qu'identifier le nombre de ménages avec enfant(s) situés à plus de 5km d'une école. Ces analyses de distances auraient pu être exécutées en vectoriel, toutefois elles auraient nécessité beaucoup plus de traitements (cf. section 4.2.1.2)

5.2.2 Les analyses obtenues par traitements au niveau des mesures

Les analyses matricielles obtenues par traitements peuvent également être exploitées en **mesure** dans un cube. Pour les mesures, il est possible de précalculer les analyses ou de les effectuer à la volée en utilisant des mesures calculées. Les analyses matricielles sont réalisées à l'aide d'opérateurs simples (somme, moyenne, min, max, >, <, etc.) appliqués de manière locale, focale ou zonale sur les valeurs des cellules d'une ou de plusieurs couches matricielles (cf. section 2.3.3). Le choix de précalculer une analyse et d'en faire une mesure ou de la réaliser à la volée dans l'outil dépend en grande partie de la complexité de cette analyse. Certaines analyses sont simples et ne nécessitent qu'une seule opération comme la somme de plusieurs couches tandis que d'autres plus complexes en nécessitent plusieurs, comme la détermination du chemin optimal. Les analyses impliquant une seule opération sont très rapides à exécuter et peuvent ainsi être réalisées de manière interactive dans un cube. Les analyses qui nécessitent plusieurs opérations engendrent des temps de traitements importants et gagnent donc à être précalculées. Une analyse matricielle précalculée peut être exploitée en mesure dans un cube matriciel ou dans un cube d'objets qui implante le matriciel dans ses dimensions spatiales. Les analyses de distances de l'expérimentation 1 (cf. section 4.2.1.2) et les analyses de coûts de l'expérimentation 2 (cf. section 4.2.2.2) sont des exemples d'analyses précalculées exploitées en mesure. Comme le montre la figure 5-1, les analyses précalculées obtenues par traitements au niveau des mesures sont également possibles en vectoriel.

5.2.3 Les analyses obtenues par traitements au niveau des mesures à la volée

La réalisation d'une **analyse matricielle à la volée** est, de son côté, possible uniquement dans un cube matriciel et s'effectue à l'aide d'une mesure calculée. Une mesure calculée est obtenue en effectuant une opération sur une ou plusieurs mesures existantes [Spofford et *al.*, 2006]. Les faits détaillés d'un cube matriciel étant des cellules d'une grille matricielle, les opérations ainsi effectuées sur les mesures existantes du cube peuvent être des opérations de Map Algebra permettant la réalisation d'analyse spatiale matricielle. Les mesures existantes sont utilisées comme couches d'entrées pour les opérations et la mesure calculée comme couche résultante (ex. Mesure calculée (taux de natalité) = mesure1 (nombre de naissances) / mesure 2 (population totale)). Les fonctionnalités de mesures calculées de l'outil JMap Spatial OLAP permettent actuellement de réaliser certaines analyses matricielles à la volée. Ces fonctionnalités permettent d'effectuer des opérations exploitant les mesures d'un fait pour calculer une nouvelle valeur de mesure pour ce même fait. Les faits détaillés d'un cube matriciel étant des cellules d'une grille matricielle, une opération effectuée sur chaque cellule d'une grille sans tenir compte de ses voisines est appelée une analyse locale. Les fonctionnalités de mesures calculées de l'outil JMap Spatial OLAP permettent ainsi la réalisation d'analyses matricielles locales à la volée dans un cube matriciel. Dans l'expérimentation 2 (cf. section 4.2.2.2), la mesure *coût global* est un exemple d'analyse locale présentement réalisable à l'aide de l'outil JMap Spatial OLAP. Le coût global pour chacune des cellules a simplement été obtenu en effectuant une somme des mesures existantes de coûts (*Coût occupation du sol, Coût pente et Coût infrastructure*). La figure 5-3 illustre la réalisation d'une telle opération dans une table de faits d'un cube matriciel et son équivalent dans un modèle matriciel.

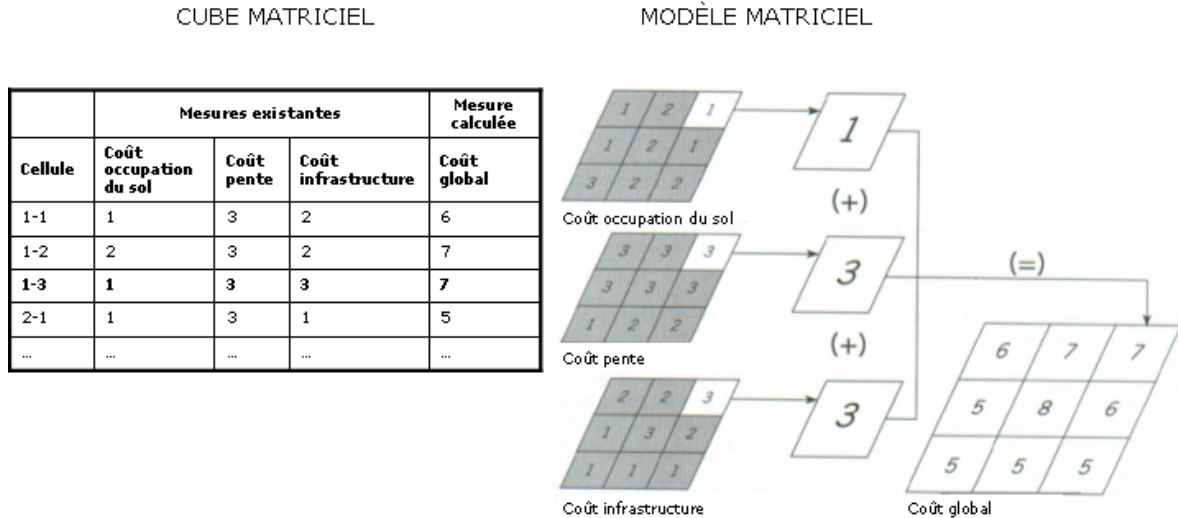


Figure 5-3 Analyse locale de Map Algebra effectuée dans un cube matriciel et dans un modèle matriciel

Des analyses simples focales et zonales sont également réalisables à la volée dans un cube matriciel. Cependant, des opérateurs doivent être implantés dans l'outil SOLAP pour permettre leur réalisation. Par exemple, dans l'expérimentation 3 (cf. section 4.2.3), nous avons implanté un opérateur dans le serveur OLAP Mondrian permettant la réalisation d'une analyse focale (Focal Rating). Cette dernière permettait la réalisation d'une analyse d'adjacence à la volée. La réalisation d'analyses à la volée n'était pas possible avec la structure vectorielle, l'implantation de la structure matricielle permet ainsi d'améliorer grandement le potentiel des outils SOLAP pour la réalisation d'analyses spatiales interactives.

En plus, s'il y a présence d'une dimension spatiale géométrique matricielle dans le cube, le résultat d'une analyse matricielle exploité en mesure peut être agrégé sur un maillage de résolution plus grossière. La simple utilisation des opérateurs d'agrégation OLAP (ex. somme, moyenne, min, max, médiane), tel qu'illustré à la figure 5-4, permet ainsi de créer des opérateurs d'analyses matricielles agrégatifs. Les opérateurs d'analyses matricielles sont des opérateurs non-agrégatifs, c'est-à-dire qu'ils retournent une nouvelle valeur pour chacune des cellules de la grille, sans modifier la résolution de cette

dernière. Toutefois, l'exploitation d'une analyse matricielle dans une mesure d'un cube combiné à l'implantation d'une dimension spatiale géométrique matricielle et à l'utilisation d'opérateurs OLAP permet la création d'analyses agrégatives.

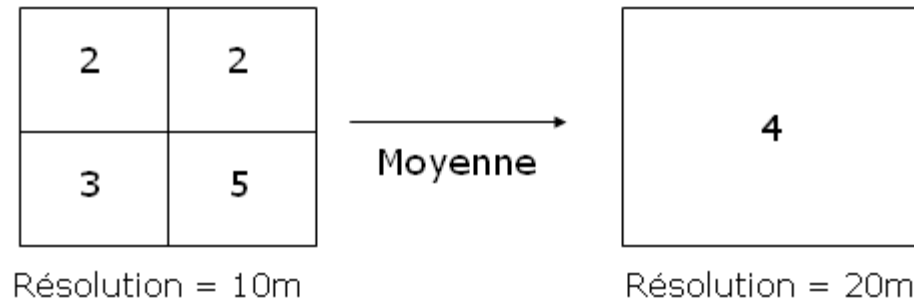


Figure 5-4 Opération d'agrégation

5.2.4 Les analyses obtenues par simple navigation dans le cube

En plus des analyses obtenues par traitements, l'implantation du matriciel dans un cube permet également la réalisation de certaines analyses **par simple navigation dans le cube** ce qui n'était pas possible avec la structure vectorielle. En effet, l'implantation du matriciel dans les dimensions d'un cube permet par simple navigation d'effectuer des analyses telles que des intersections spatiales et des analyses de superposition spatiale. Les outils SOLAP actuels permettent la visualisation d'une seule dimension spatiale géométrique à la fois. Il est donc impossible de représenter cartographiquement le résultat d'une analyse à l'intersection entre deux découpages vectoriels tels que le découpage administratif et le découpage de la santé sans avoir précalculé et stocké l'ensemble des intersections possibles entre ces deux découpages. Cette procédure peut être très fastidieuse si les géométries des découpages vectoriels sont complexes. L'implantation du matriciel dans les dimensions spatiales d'un cube permet de faire le lien entre les deux dimensions géométriques vectorielles en représentant l'intersection

spatiale sans devoir stocker la géométrie de l'ensemble des intersections possibles. La réalisation d'une analyse d'intersection spatiale entre deux découpages vectoriels dans un cube nécessite soit l'implantation d'une dimension spatiale géométrique matricielle soit l'implantation d'un niveau fin matriciel dans l'une des dimensions dont on veut faire l'intersection (ce qui conduit à la création d'une dimension géométrique hybride). La figure 5-5 illustre les possibilités d'implantation du matriciel dans les dimensions d'un cube permettant la réalisation de l'intersection spatiale entre deux découpages vectoriels (découpage administratif et découpage santé).

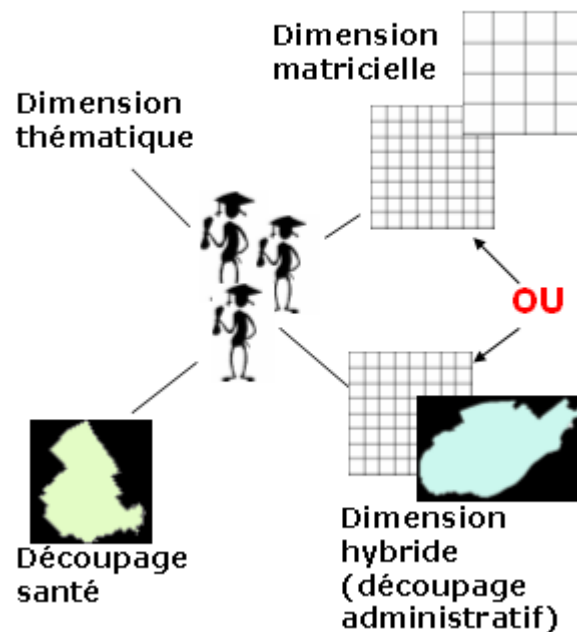


Figure 5-5 Possibilités d'implantation du matriciel dans les dimensions spatiales permettant la visualisation de l'intersection spatiale entre deux découpages vectoriels (un niveau fin matriciel pourrait également être implanté dans la dimension géométrique vectorielle découpage santé).

Le résultat d'une analyse à l'intersection entre un ou plusieurs membres de la dimension découpage santé et un ou plusieurs membres de la dimension découpage administratif peut être représenté cartographiquement en utilisant soit la dimension géométrique matricielle ou la dimension découpage

administratif au niveau fin matriciel. Le résultat de l'intersection est représenté sous forme matricielle tel qu'illustré à la figure 5-6.

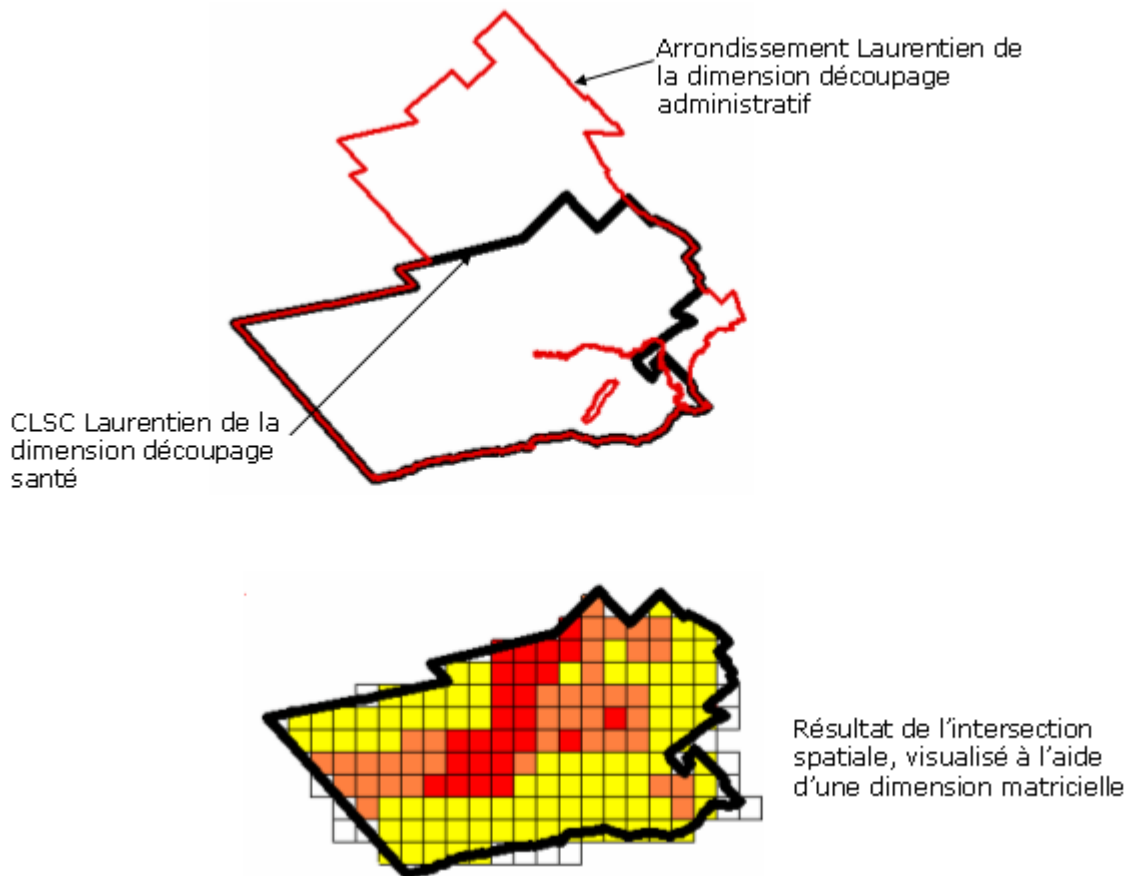


Figure 5-6 Intersection spatiale entre un membre de la dimension découpage administratif et un membre de la dimension découpage santé.

La réalisation d'une intersection spatiale est possible dans les deux types de cubes, d'objets ou matriciel. Toutefois, pour le cube d'objets, la couverture spatiale des données doit être complète pour pouvoir représenter l'intersection spatiale entre deux membres vectoriels.

Un deuxième type d'analyse réalisable par simple navigation dans les dimensions d'un cube est la superposition spatiale (Spatial Overlay). Dans un cube matriciel, les faits détaillés étant des cellules d'une grille matricielle et les dimensions, des caractéristiques de ces cellules, la simple navigation dans les dimensions d'un cube matriciel permet la réalisation d'une grande quantité de

superpositions spatiales. Le croisement entre deux dimensions thématiques d'un cube matriciel correspond à superposer deux couches matricielles dans un modèle matriciel. Il est ainsi possible de croiser une ou plusieurs dimensions d'un cube matriciel en sélectionnant un ou plusieurs membres de ces dimensions et de visualiser le résultat de la superposition par l'utilisation du matriciel dans les dimensions spatiales. Dans l'expérimentation 2 (cf. section 4.2.2.2), les dimensions occupation du sol et pente permettent par exemple d'effectuer des analyses telles qu'identifier les endroits en « zone urbaine » où la pente est « >7% ». La réalisation d'une telle analyse correspond à superposer une couche d'occupation du sol et une couche de pente pour lesquelles le type d'occupation du sol « urbain » et la classe de pente « >7% » sont bien identifiés. Dans le SOLAP, la réalisation de cette analyse s'effectue très simplement par la sélection du membre « >7% » de la dimension pente et le membre « urbain » de la dimension occupation du sol. Il suffit ensuite d'afficher sur une carte une mesure arbitraire telle que le nombre de cellules par le biais d'une dimension géométrique matricielle ou hybride. La figure 5-7 illustre cette analyse réalisée dans l'outil JMap Spatial OLAP.

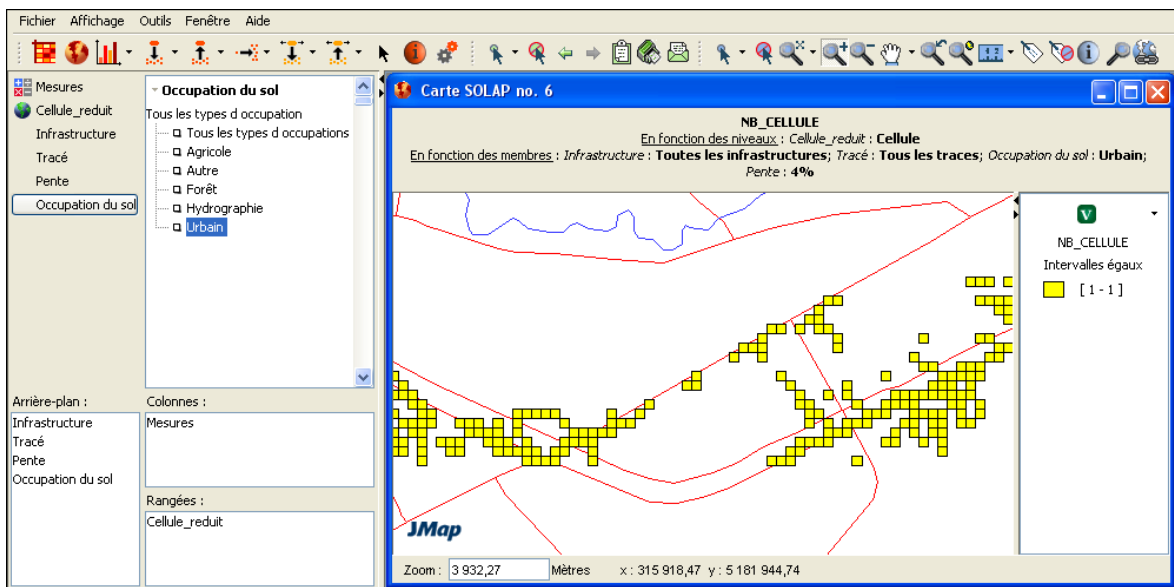


Figure 5-7 Analyse de superposition spatiale dans un cube matriciel

5.3 Conclusion

Nous avons dans ce chapitre initié l'étude du potentiel de la structure matricielle pour l'analyse spatiale dans les outils SOLAP. Ceci nous a permis d'identifier et de mettre en avant les différentes possibilités d'exploitation des analyses matricielles dans un cube. Les analyses matricielles peuvent être soit obtenues par traitements ou par simple navigation dans les dimensions d'un cube. Pour les analyses obtenues par traitements, il est possible de les exploiter dans les dimensions ou dans les mesures en mode précalculé ou à la volée. Pour les analyses obtenues par navigation dans le cube, il est entre autres possible de réaliser des intersections spatiales et des analyses de superposition spatiale. Pour ces analyses, les possibilités sont directement liées à la conception du cube.

Les analyses matricielles peuvent être exploitées autant dans un cube matriciel que dans un cube d'objets qui implante le matriciel dans ses dimensions spatiales. Toutefois, les analyses réalisées à la volée et exploitées en mesure sont possibles uniquement dans le cube matriciel. Pour les analyses obtenues par navigation, il est possible de les réaliser dans un cube d'objets, mais la couverture spatiale des données doit être complète sur l'ensemble du territoire couvert.

Afin d'illustrer le potentiel du matriciel pour améliorer les analyses spatiales dans les outils SOLAP, nous présentons dans le tableau 5-1 les possibilités d'exploitation d'analyses vectorielles ou matricielles en fonction des types de cubes avec ou sans l'implantation du matriciel. On peut voir dans ce tableau, que l'intégration du matriciel apporte d'avantages de possibilités d'exploitation des analyses spatiales dans un cube.

L'identification des différentes possibilités d'exploitation des analyses matricielles a permis de démontrer le potentiel du matriciel à réaliser plusieurs analyses spatiales dans les outils SOLAP. La simplicité et la rapidité des opérateurs d'analyses matricielles permettent même la réalisation de certaines

analyses à la volée. Les analyses présentement réalisables à la volée à l'aide des fonctionnalités existantes de l'outil JMap Spatial OLAP sont des analyses de type local. L'implantation de nouveaux opérateurs permettrait la réalisation de plusieurs autres analyses telles que des analyses focales ou zonales, il serait ainsi possible de réaliser à la volée de nouvelles analyses spatiales matricielles dans un SOLAP. Dans le cadre de cette recherche, nous avons testé une seule analyse spatiale matricielle à la volée ce qui a été suffisant pour en démontrer la faisabilité. Pour les autres analyses, nous posons néanmoins, l'hypothèse que leur l'exécution ne serait pas plus lente dans un système tel que le SOLAP que dans un système transactionnel.

Type de cube \ Type d'exploitation	Par navigation	Par traitement(s)		
		Dimension	Mesure	
			Précalculée	À la volée
Cube d'objets sans matriciel		X	X	
Cube d'objets avec matriciel en dimension	X*	X	X	
Cube matriciel	X	X	X	X

* Possible seulement si couverture complète des données

Tableau 5-1 Exploitation des analyses spatiales en fonction du type de cube

CONCLUSION

L'objectif principal de cette recherche consistait à définir un cadre théorique permettant l'intégration de la structure matricielle dans les cubes spatiaux. Cet objectif a été atteint par l'accomplissement de trois objectifs secondaires. Le premier de ces objectifs consistait à décrire les façons possibles d'intégrer la structure matricielle dans un cube de données. Nous avons répondu à cet objectif dans le chapitre 3 en proposant deux approches visant à intégrer l'aspect matriciel dans les cubes de données. La première approche propose d'intégrer le matriciel dans les dimensions spatiales et la seconde propose de l'intégrer dans les faits. L'intégration du matriciel dans les dimensions spatiales a nécessité la définition de quatre nouvelles dimensions spatiales soit : la dimension géométrique matricielle, la dimension géométrique hybride, la dimension mixte matricielle et la dimension mixte hybride. L'intégration du matriciel dans les faits a ensuite nécessité la définition de deux types de cubes spatiaux soit : le cube d'objets et le cube matriciel. Ces concepts peuvent être combinés tel que toute combinaison de type de cube et de type de dimension est permise. La structure matricielle peut ainsi être intégrée à la fois dans les dimensions d'un cube et dans ses faits. Les concepts présentés dans le chapitre 3 constituent les résultats théoriques de cette recherche. Ces résultats ont permis d'enrichir les concepts SOLAP et ont contribué directement à l'avancement des connaissances en géomatique décisionnelle. Ce qui est présenté de ce mémoire constitue le résultat final, l'ensemble des réflexions menant à ce résultat non toutefois pas été présentées.

Le deuxième objectif secondaire de cette recherche consistait à démontrer la faisabilité d'exploiter la structure matricielle dans les cubes de données. Nous avons répondu à cet objectif au chapitre 4 par la réalisation de trois expérimentations exploratoires et d'une expérimentation formelle. Ces expérimentations ont non seulement permis de démontrer la faisabilité d'exploiter la structure matricielle dans un cube de données, mais ont

également permis de tester et de valider les concepts théoriques qui avaient été définis précédemment. Les expérimentations exploratoires ont dans un premier temps permis de tester les concepts élaborés et l'expérimentation formelle a ensuite validé l'utilité d'utiliser le matriciel dans un cube de données en répondant à un besoin réel et actuel.

Le troisième et dernier objectif secondaire de cette recherche consistait à initier l'étude du potentiel d'analyse spatiale matricielle en identifiant des possibilités d'exploitation de ces analyses dans les outils SOLAP. Nous avons initié cette étude dans le chapitre 5 par l'identification des possibilités d'exploitation d'analyses spatiales matricielles dans les outils SOLAP en fonction du type de cube utilisé et par la démonstration, à l'aide d'exemples, de la réalisation de certaines de ces analyses. Dans un cube, les résultats d'analyses spatiales matricielles peuvent être obtenus soit par traitements, soit par navigation dans le cube. Pour les analyses obtenues par traitements, il est possible de les précalculer ou de les exécuter à la volée si le cube utilisé est un cube matriciel. L'étude du potentiel d'analyse spatiale matricielle réalisée au chapitre 5 n'est pas exhaustive, c'est-à-dire qu'elle repose uniquement sur les expérimentations réalisées au chapitre 4, mais elle permet néanmoins de démontrer le potentiel réel du matriciel à enrichir les analyses spatiales dans les outils SOLAP.

La réalisation des différentes expérimentations et l'étude du potentiel pour l'analyse spatiale ont également permis d'identifier les principaux avantages et inconvénients d'utiliser la structure matricielle par rapport à une structure vectorielle dans des cubes spatiaux. Les avantages et inconvénients ainsi identifiés sont :

Avantages :

- le matriciel est mieux adapté pour le traitement des phénomènes à variation continue et l'analyse de surface;
- il est mieux adapté pour la gestion de plusieurs découpages complexes (ex. superposition des découpages administratifs et des zones vertes);

- il permet de mettre l'information en densité automatiquement. Puisque les cellules d'une grille matricielle sont de mêmes tailles, des données telles que la population peuvent être exprimées et comparées sans nécessiter des calculs de densité;
- il libère l'utilisateur des contraintes imposées par les découpages administratifs ou autres qui sont prédéfinis;
- il permet d'exploiter des résultats précalculés d'analyses spatiales matricielles;
- il permet de faire certaines analyses spatiales à la volée (cas du cube matriciel) ce qui n'est pas possible en vectoriel dû à la complexité des analyses;
- et finalement, il permet de faire d'autres regroupements (ex. le long du St-Laurent) que ceux prédéfinis par les hiérarchies (ex. découpage administratif)

Ce dernier point constitue un avantage majeur de l'intégration de la structure matricielle dans les cubes de données. En effet, l'analyse d'un phénomène spatial (ex. la population atteinte d'une maladie grave) selon un découpage prédéfini (ex. le découpage administratif) permet de comparer la distribution spatiale de ce phénomène sur les différents membres du découpage. Toutefois, il ne permet pas de voir la distribution spatiale réelle de ce phénomène. Puisque le matriciel est un découpage régulier, il permet de voir cette distribution et de potentiellement faire d'autres regroupements que ceux prédéfinis par la hiérarchie (ex. découvrir que les gens atteints d'une maladie résident à proximité d'une ligne à haute tension).

Inconvénients :

- l'utilisation du matriciel dans les dimensions spatiales occasionne une augmentation du nombre de membres;
- l'utilisation du matriciel dans les faits produit une augmentation du nombre de faits;
- l'utilisation du matriciel nécessite souvent la gestion de relations multiples entre les niveaux des dimensions spatiales ou entre la table de faits et les dimensions.
- et finalement, le matriciel représente l'information de manière plus grossière que le vectoriel.

L'ensemble de cette recherche a permis de faire un premier pas pour l'utilisation de la structure matricielle dans les cubes spatiaux. Nous avons défini le cadre théorique permettant l'intégration de données matricielles dans les cubes, nous l'avons testé et validé dans le cadre d'expérimentations et nous avons finalement initié le potentiel du matriciel à améliorer les analyses spatiales dans les outils SOLAP. Maintenant que ce premier pas est fait, nous proposons dans la section suivante des travaux futurs qui permettront d'enrichir l'exploitation de la structure matricielle dans les outils SOLAP.

Travaux futurs

L'ensemble des expérimentations ont permis de tester et valider les concepts développés, toutefois ces expérimentations ont été réalisées à l'aide d'une « simulation matricielle » plutôt qu'une structure matricielle pure. Puisque l'outil JMap Spatial OLAP ne supporte actuellement pas de structure matricielle en tant que telle pour les analyses multidimensionnelles, l'utilisation d'une « simulation matricielle » nous a permis de tester les concepts élaborés tout en étant supportée par l'outil JMap Spatial OLAP actuel. Toutefois, il serait intéressant d'étudier la possibilité d'y implanter une structure matricielle pure. L'implantation d'une structure matricielle pure ou potentiellement d'une autre structure en tessellation régulière permettrait d'accélérer les traitements, ce qui améliorerait potentiellement les analyses spatiales interactives. Des travaux de recherche dans le cadre du projet avec la défense canadienne ont déjà été initiés en ce sens avec la technologie Pyxis (<http://www.pyxisinnovation.com/>) afin d'identifier son potentiel pour la gestion de cubes en tessellation hexagonale.

Maintenant qu'il est possible de construire des cubes de données supportant des données matricielles, il serait éventuellement intéressant de définir les fonctionnalités d'un SOLAP matriciel. Afin de faciliter la navigation dans le SOLAP, des opérateurs spécifiques à l'exploration interactive du contenu matriciel pourraient être développés. Ces potentiels opérateurs de navigation

seraient propres à un cube qui contient des données matricielles et pourraient inclure des opérateurs de forage comme un « drill-down to raster ». Cet opérateur permettrait par exemple de naviguer d'un niveau vectoriel à un niveau matriciel dans une dimension géométrique hybride. Dans le même ordre d'idées, d'autres opérateurs tels que « drill-up to vector », « drill-up to raster », « drill-across to raster », « drill-across to vector », « drill-zonal to raster », etc. pourraient être développés. De plus, l'une des limites de l'outil actuel réside dans le rafraîchissement de l'arbre des dimensions impliquant plusieurs membres localisés sous différentes branches. Pour ces cas, le rafraîchissement ralentit énormément les processus et c'est souvent le cas pour les requêtes impliquant l'ouverture de l'arbre d'une dimension matricielle. En ce sens, de nouvelles fonctionnalités pourraient cacher l'arbre de dimension lorsque l'information sur les cellules n'est pas significative (ex. numéro de ligne et de colonne).

Dans le chapitre 5, nous avons initié l'étude du potentiel de la structure matricielle pour l'analyse spatiale dans les outils SOLAP. Nous avons vu qu'il était actuellement possible de réaliser des analyses de type local à la volée à l'aide des fonctionnalités existantes de l'outil JMap Spatial OLAP. Des travaux futurs pourraient poursuivre cette étude en développant de nouveaux opérateurs d'analyses matricielles interactives. L'implantation de nouveaux opérateurs permettrait la réalisation de plusieurs autres analyses telles que des analyses focales ou zonales. Il serait ainsi possible de réaliser à la volée une foule d'analyses spatiales matricielles dans un SOLAP. Les résultats d'analyses pourraient ensuite être utilisés pour former de nouveaux membres spatiaux. C'est-à-dire qu'il serait potentiellement possible de créer de nouveaux membres formés de regroupements de cellules qui correspondent à un résultat d'analyse. Ce processus correspondrait à créer des hiérarchies de dimensions spatiales à la volée.

Et finalement, afin d'aider les utilisateurs à construire des cubes de données selon leurs besoins et leurs données, il serait intéressant de construire un outil d'aide à la conception de cubes. Cet outil permettrait de spécifier le type de

cube à construire et le type de dimensions nécessaires pour répondre aux besoins des utilisateurs.

Bibliographie

- AHMED, T. (2006), *Continuité spatiotemporelle dans les entrepôts de données et les modèles multidimensionnels*, Thèse présentée et soutenue pour l'obtention du grade de Docteur, Université Claude Bernard Lyon I, 182p.
- ARMSTRONG, Marc. P., DENSHAM, Paul.J, *Towards a network map algebra*, http://www.uiowa.edu/~geog/faculty/armstrong/Net_Map Alg_Paper.pdf
- *BÉDARD, Y., MERRETT, T., HAN, J. (2008), «Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery», In: *Geographic Data Mining and Knowledge Discovery*, 2e édition, chap.3, Taylor & Francis.
- *BÉDARD, Y., RIVEST, S., PROULX, M.-J. (2007), « Spatial Online Analytical Processing (SOLAP) : Concepts, Architectures, and Solutions from a Geomatics Engineering Perspective », Chapter in : *Data Warehouses and OLAP, Concepts, Architectures and Solutions*, Wrembel et Koncilia, IRM Press, 332p.
- *BÉDARD, Y., PROULX, M.-J., RIVEST, S. (2006a), *Notions de modélisation multidimensionnelle (conceptuelle)*, Notes de cours SCG-66124, Département des Sciences Géomatiques, Université Laval.
- BÉDARD, Y., PROULX, M.J. et RIVEST, S. (2005), «Enrichissement du OLAP pour l'analyse géographique : exemples de réalisation et différentes possibilités technologiques», *Revue des Nouvelles Technologies de l'Information - Entrepôts de données et l'Analyse en ligne*, sous la direction de F. Bentayeb, O. Boussaïd, J. Darmont et S. Loudcher, France, Cépaduès-Éditions, p. 1-20.
- *BÉDARD, Yvan (2004), *Amélioration des capacités décisionnelles des SIG par l'ajout d'un module SOLAP (Spatial On-Line Analytical Processing)*, Université Aix-Marseille, École Polytechnique Universitaire de Marseille, Filière Génie Industriel et Informatique, 8 avril (Professeur invité).
- *BÉDARD, Yvan (2003), *Notes de cours Fondements des SIG et SIG et analyse spatiale*. Département des Sciences Géomatiques, Université Laval, Québec.
- *BÉDARD, Y., MERRETT, T., HAN, J. (2001), «Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery», Chapter of the book *Geographic Data Mining and Knowledge Discovery*, edited by H. Miller and J. Han, Research Monographs in GIS series edited by Peter Fisher and Jonathan Raper, Taylor & Francis.

- *BÉDARD, Y., LARRIVÉE, S., PROULX, M-J., CARON, P-Y., et LÉTOURNEAU, F. (1997), *Geospatial Data Warehousing : Positionnement technologique et stratégique*. Rapport pour le Centre de recherche pour la défense de Valcartier (CRDV), 79p.
- BÉDARD, Yvan (1991), « Les logiciels SIG : Évolution via l'intégration de données multisources », *Journal de la Société Française de Photogrammétrie et de Télédétection*, No. 122, p. 58-63.
- BELL, S. B.M., DIAZ, B.M., HOLROYD, F., and JACKSON, M.J. (1983), «Spatially referenced methods of processing raster and vector data», *Image and vision computing*, Vol. 1, No. 4, p. 211-220.
- BERNHARDBSEN, T. (2002), *Geographic Information Systems An Introduction*, Third Edition, John Wiley and Sons, INC., New-York.
- BIMONTE, S., TCHOUNIKINE, A., MIQUEL, M., LAURINI, R. (2007), « Vers l'intégration de l'analyse spatiale et multidimensionnelle », Acte de conférence, Colloque SAGEO 2007, Clermont-Ferrand.
- *BONHAM-CARTER, Graeme F. (1994), *Geographic information systems for geoscientists: Modelling with GIS*. Canada, Pergamon, Computer methods in the geosciences, vol. 13, 398p.
- *BURROUGH, Peter A., MCDONNELL, Rachael A. (1998), *Principles of geographical information systems*. Oxford University Press, 333p.
- *CAMARA, G., PALOMO, D., CARAXO, R., OLIVEIRA, O. (2005) « Towards a generalized map algebra : principles and data types », in VII Workshop Brasileiro de Geoinformática. 2005. Campos do Jordão: SBC.
- CALOZ, R., COLLET, C. (1997), *Précis de télédétection, traitements numériques d'images de télédétection*, Presse de l'université du Québec, Québec, 474p.
- *CARRON, Pierre-Yves (1998), *Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle*. Mémoire présenté à la Faculté des études supérieures de l'Université Laval, Québec, 129 p.
- CHAMPOUX, Pierrette (1991), *Étude sur les fonctions d'analyse spatiale à utiliser dans un SIRS appliqué à l'exploration minière*. Mémoire présenté à la Faculté des études supérieures de l'Université Laval, Québec.
- CHANG, Kang-Tsung (2004), *Introduction to geographic information systems* 2nd edition, New-York, McGraw-Hill Higher Education, 400p.
- *CHRISMAN, N. (1975), « Topological information systems for geographic representation », Proceedings of second International Symposium on

Computer-Assisted Cartography (Auto-carto 2), Falls Church: ASPRS/ACSM, p. 346-351.

- *CHRISMAN, N. (1978), «Concepts of space as a guide to cartographic data structures », in Dutton, G.(Ed.), Proceedings of the first International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Cambridge, MA: Harvard Laboratory for Computer Graphics and Spatial Analysis, p. 1-19.
- *CHRISMAN, Nicholas (1997), *Exploring geographic information systems*. New York, John Wiley & Sons Inc, 298p.
- *CODD, E. F. *et al.* (1993), *Providing OLAP to User-Analysts : An IT Mandate*, An Hyperion Solutions White Paper, Hyperion Solutions, 24 p.
- COLLET, Claude (2005), «Analyse spatiale, géomatique et systèmes d'information géographique », *Revue internationale de géomatique*, Vol. 15, No.4, p. 393-414.
- *COLLET, Claude (1992), *Systèmes d'information géographique en mode image*. Lausanne, Presses polytechniques et universitaires romandes, coll. Gérer l'environnement, 186p.
- CROMLEY, R.G., HANINK, D.M. (1999), «Coupling land use allocation models with raster GIS», *Journal of geographical systems*, Vol 1, No 2, pp. 137-153.
- *DEMERS, Michael N. (2002), *GIS modeling in raster*. United States, John Wiley & sons, Inc., 203p.
- DEURSEN, Van, W.P.A. (1995), *geographical information systems and dynamic models: development and application of a prototype spatial modelling language*. PhD Thesis, Faculty of Spatial Sciences University of Utrecht, Netherlands.
- *DONNAY, Jean-Paul (2005), « Formalisation des informations géographiques en mode maillé », *Revue internationale de géomatique*, Vol. 15, No.4, p. 415-438.
- EGENHOFER, M.J. (1989), « A formal definition of binary topological relationships », *Lecture Notes in Computer Science*, Vol. 367, p.457-472.
- FRANKLIN, C. (1992). «An Introduction to Geographic Information Systems: Linking Maps to Databases», *Database*, April, pp. 13-21
- FORTIN, M-J., PAYETTE, S. (2002), « How to test the significance of the relation between spatially autocorrelated data at the landscape scale: A

case study using fire and forest maps », *Ecoscience*, Vol. 9, No. 2, p.213.

FOTHERINGHAM, S., ROGERSON, P. (1994), *Spatial Analysis and GIS*. Taylor and Francis.

GIS WORLD (1989), Special Report, GIS Technology 89 : Results of the 1989, GIS WORLD Geographic Information Systems Survey.

GOODCHILD, M.F., et all (2007), «Towards a general theory of geographic representation in GIS», *International Journal of Geographical Information Science*, Vol.21, No.3, p. 239-260.

GOODCHILD, M.F (2005), «Geographic information systems», *Encyclopedia of Social Measurement 2*, p.107-113.

GOODCHILD, M.F. (2001), «Spatial Analysis and GIS», 2001 Esri user conference, Pre-Conference Seminar.

*GOUVERNEMENT DU QUÉBEC (2007), *Enquête OD de 2001*, http://www.mtq.gouv.qc.ca/portal/page/portal/ministere/ministere/recherche_innovation/modelisation_systemes_transport/enquetes_origine_destination/quebec/enquete_2001, Transports Québec, site consulté le 29 avril 2008.

*HAN, J., STEFANOVIC, N. and KOPERSKI, K. (1998), «Selective Materialization: An Efficient Method for Spatial Data Cube Construction», *Proc. Pacific-Asia Conf. on Knowledge Discovery and Data Mining*, Melbourne, Australia, p. 144-158.

HORNER, J., et al. (2004), « An Analysis of Additivity in OLAP Systems », *Proceedings of the 7th ACM international Workshop on Data Warehousing and OLAP (DOLAP '04)*, ACM Press, pp. 83-91.

*ISO 19123:2005(E) (2005), *Information géographique—Schéma de la géométrie et des fonctions de couverture*. First Édition.

JOERIN, Florent (1997), *Décider sur le territoire Proposition d'une approche par utilisation de SIG et de méthode d'analyse multicritère*. Thèse présentée au département de génie rural de l'école polytechnique fédérale de Lausanne, Lausanne EPFL.

*KIMBALL, R. and ROSS, M. (2002), *The Data Warehouse Toolkit Second Edition : The complete guide to dimensional modeling*. John Wiley and Sons, Inc., 436p.

LATOURET, Philippe, *Spatialist Analyse spatiale et géomarketing*, <http://www.spatialist.fr/>, site consulté le 21 mars 2007.

- *LAURINI, Robert, THOMPSON, Derek (1992), *Fundamentals of spatial information systems*. London, The A.P.I.C Series number 37, 680p.
- *LETERRIER, E., PIROT, F., SAINT-GERAND, T. (1988), «La représentation cartographique des phénomènes continus dans l'espace géographique», *MappeMonde*, No. 2/88.
- *LÉVESQUE, M.-A., BÉDARD, Y., GERVAIS, M., DEVILLERS, R. (2007), « Towards Managing the Risks of Data Misuse for Spatial Datacubes ». International Symposium on Spatial Data Quality. ITC, June 13-14, Enschede, The Netherlands.
- MALINOWSKI, E., ZIMANYI, E. (2005), « Spatial Hierarchies and Topological Relationships in the Spatial MulriDimER model », In Proc. Of the 22nd British Nat. Conf. on Databases, BNCOD22, number 3567 in LNCS, p. 17-28, Sunderland, UK.
- MARCHAND, P., BRISEBOIS, A., BÉDARD, Y., EDWARDS, G. (2003), «Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis», *Workshop ISPRS*, October 2-3, Québec, Canada.
- *MARCHAND, Pierre (2004), *The Spatio-Temporal Topological Operator Dimension, a Hyperstructure for Multidimensional Spatio-Temporal Exploration and Analysis*. Mémoire présenté à la Faculté des études supérieures de l'Université Laval, Québec, 108 p.
- *MCHUGH, R., BILODEAU, F., RIVEST, S., BÉDARD, Y. (2006), « Analyse du potentiel d'une application SOLAP pour une gestion efficace de l'érosion des berges en Gaspésie Îles-de-la-Madeleine », *Géomatique 2006*, 25-26 octobre 2006, Montréal, Canada.
- MENNIS, Jeremy, *Multidimensionnal map algebra*, http://igre.emich.edu/geocomputation2005/abstract_list/1000009mennis.pdf. site consulté le 20 juin 2007.
- *MIQUEL, M., BÉDARD, Y. et BRISEBOIS, A. (2002), « Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie », *Ingénierie des Systèmes d'information*, vol. 7, No. 3, p. 89-111.
- MOLENAAR, Martien (1998), *An introduction to the theory of spatial object modelling for GIS*, London, Taylor & Francis, 246p.
- *NEWELL, A. (1990), *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

- NORMAND, Pierre (1999), *Modélisation des contraintes d'intégrité spatiale, théorie et exemples d'applications*. Essai présenté à la Faculté des études supérieures de l'Université Laval, Québec, 95 p.
- *NORMAND Pierre *et al.* (2002), « Chapitre 6 – Modélisation des contraintes d'intégrité spatiale », Recueil de lecture pour le cours GMT-20843 Fondements des SIG, Université Laval, Session Automne 2003, 13 p.
- *OQLF (2004), *Office québécois de la langue française*, <http://www.oqlf.gouv.qc.ca/>.
- *OLAP Council (1995), *OLAP and OLAP Server Definitions*, <http://www.olapcouncil.org/research/resrchly.htm>, site consulté le 12 juin 2007.
- OGC OPEN GIS CONSORTIUM INC (1999), *OpenGIS Simple Features Specification For SQL*, Revision 1.1, OpenGIS Project Document 99-049.
- *OGC OPEN GIS CONSORTIUM INC (2006), *The OpenGIS Abstract Specification -Topic 6: Schema for coverage geometry and functions*, OpenGIS Project Document 99-049, Version 7.0, 75p.
- ORACLE SPATIAL (2005), *User's guide and reference*, 10g release 2, <http://www.oracle.com/technology/documentation/spatial.html>, site consulté le 12 juin 2007.
- PAPADIAS, D., KALNIS, P., ZHANG, J., TAO, Y. (2001), « Efficient OLAP operations in spatial data warehouses », 7th International symposium, SSTD 2001 (LNCS 2121), p. 443-459.
- PENDSE, Nigel (2005a), « What is OLAP? », *The OLAP Report*, <http://www.olapreport.com/fasmi.htm>, site consulté le 7 février 2008.
- PENDSE, Nigel (2005b), « Database explosion », *The OLAP Report*, <http://www.olapreport.com>, site consulté le 7 février 2008.
- *PEUQUET, D.J. (1984), « A conceptual framework and comparison of spatial data models », *Introductory Readings in Geographic Information Systems*, London: Taylor & Francis, p. 250-285.
- *PUMAIN, Denise, SAINT-JULIEN, Thérèse (2004), *L'analyse spatiale Localisations dans l'espace*. Armand Colin, Paris, 167p.
- *PROULX, M.-J., BERNIER, E., MCHUGH, R. (2008), *Amélioration de la représentation visuelle et de la gestion des découpages cartographiques dans l'application GÉOLAP*. Rapport de recherche remis à RDDC-Valcartier, Centre de recherche en géomatique, Université Laval.

- *PROULX, M.-J., RIVEST, S. (2007), *Évaluation des produits commerciaux offrant des capacités combinées d'analyse multidimensionnelle et de cartographie*. Rapport de recherche pour la Chaire de recherche industrielle CRSNG en bases de données géospatiales décisionnelles, Département des Sciences géomatiques, Université Laval.
- PROULX, Marie-Josée (2004), *Comparaison de l'approche transactionnelle des SIG avec l'approche multidimensionnelle pour l'analyse de données spatio-temporelles*, Montréal, Colloque Géomatique 2004 – Un choix stratégique!, 27-28 octobre 2004, 12 p.
- *PROULX, Marie-Josée (1995), *Développement d'un nouveau langage d'interrogation de bases de données spatio-temporelles*. Essai présenté à la Faculté des études supérieures de l'Université Laval, Québec, 148 p.
- RIVEST, S., BÉDARD, Y., PROULX, M.-J., NADEAU, M., HUBERT, F. et PASTOR, J. (2005), «SOLAP: Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data», *Journal of International Society for Photogrammetry and Remote Sensing (ISPRS)*, Vol 60, No 1, p. 17-33.
- RIVEST, S., BÉDARD, Y., PROULX, M.-J., & NADEAU, M. (2003), «SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis», Workshop ISPRS, Quebec, Canada, October 2-3, Québec, Canada.
- *RIVEST, S., BÉDARD, Y., MARCHAND, P. (2001), «Toward better support for spatial decision making: defining the characteristics of spatial on-line analytical processing (SOLAP)», *Geomatica*, vol. 55, No. 4, p. 539-555.
- *RIVEST, Sonia (2000), *Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG*. Essai présenté à la Faculté des études supérieures de l'Université Laval, Québec, 84 p.
- ROUSSET-DESCHAMPS, Marcel, *Master Géomarketing et stratégies territoriales des entreprises et des institutions publiques*, <http://www.master-geomarketing.com/>, site consulté le 21 mars 2007.
- SHEKHAR, S., ZHANG, P., HUANG, R., VATSAVAI, R. (2003a), *Trend in Spatial Data Mining, as a chapter to appear in Data Mining: Next Generation Challenges and Future Directions*, A. Joshi, K. Sivakumar, and Y. Yesha(eds.), AAAI/MIT Press.
- SHEKHAR, Shashi, CHAWLA, Sanjay (2003b), *Spatial databases A tour*. New Jersey, Pearson education inc., Prentice Hall, 262p.

- SOURIS M., (2002), *La construction d'un système d'information géographique, principes et algorithmes du système SAVANE*, Thèse de Doctorat, Université de La Rochelle, 505 p.
- *SPOFFORD, G., HARINATH, S., WEBB, C., HAI HUANG, D., CIVARDI, F. (2006), *MDX Solutions with Microsoft SQL Server Analysis Services 2005 and Hyperion Essbase*, Second edition, Wiley Publishing Inc.
- STEFANOVIC, N., HAN, J. and KOPERSKI, K. (2000), «Object-based selective materialization for efficient Implementation of spatial data cubes», *IEEE Transactions on Knowledge and Data Engineering*, Vol 12, No 6.
- *STEFANOVIC, N. (1997), *Design and implementation of On-Line Analytical processing (OLAP) of spatial data*. MSc thesis, Simon Fraser University, Canada.
- *THOMSEN, E., (2002), *OLAP Solutions: Building Multidimensional Information Systems*, 2nd edition, John Wiley and Sons.
- THOMSEN, E., SPOFFORD, G., CHASE, D. (1999), *Microsoft OLAP Solutions*, John Wiley and Sons, 495p.
- *TOMLIN, C.D. (1990), *Geographic information systems and cartographic modelling*. Prentice Hall, N.J., 249p.
- TSUI, P.H.Y, BRIMICOMBE, A.J. (1997), «The hierarchical tessellation model and its use in spatial analysis», *Transactions in GIS*, Vol.2, No.3, p.267.
- VERBYLA, D. (2002), *Practical GIS Analysis*. Taylor and Francis, London and New-York.
- VITT, E., LUCKEVIC, M., MISNER, S. (2002), *Business intelligence: Making Better decisions faster*. Microsoft Press, 1 edition, 224p.
- *WORBOYS, M.F. (1995), *GIS a Computing Perspective*. Taylor & Francis, London, 376p.
- *WORBOYS, M.F. (1994), «Object-oriented approaches to geo-referenced information», *Geographical Information Systems*, Vol.8, No.4, pp.385-399.
- SCHMIT, C., ROUNSEVELL, M.D.A., LA JEUNESSE, I. (2006), « The limitation of spatial land use data in environmental analysis », *Environmental science and policy*, Vol 9, No 2, pp. 174-188.

ANNEXE A

Schémas des cubes et diagrammes de membres des dimensions

Schémas expérimentation 1

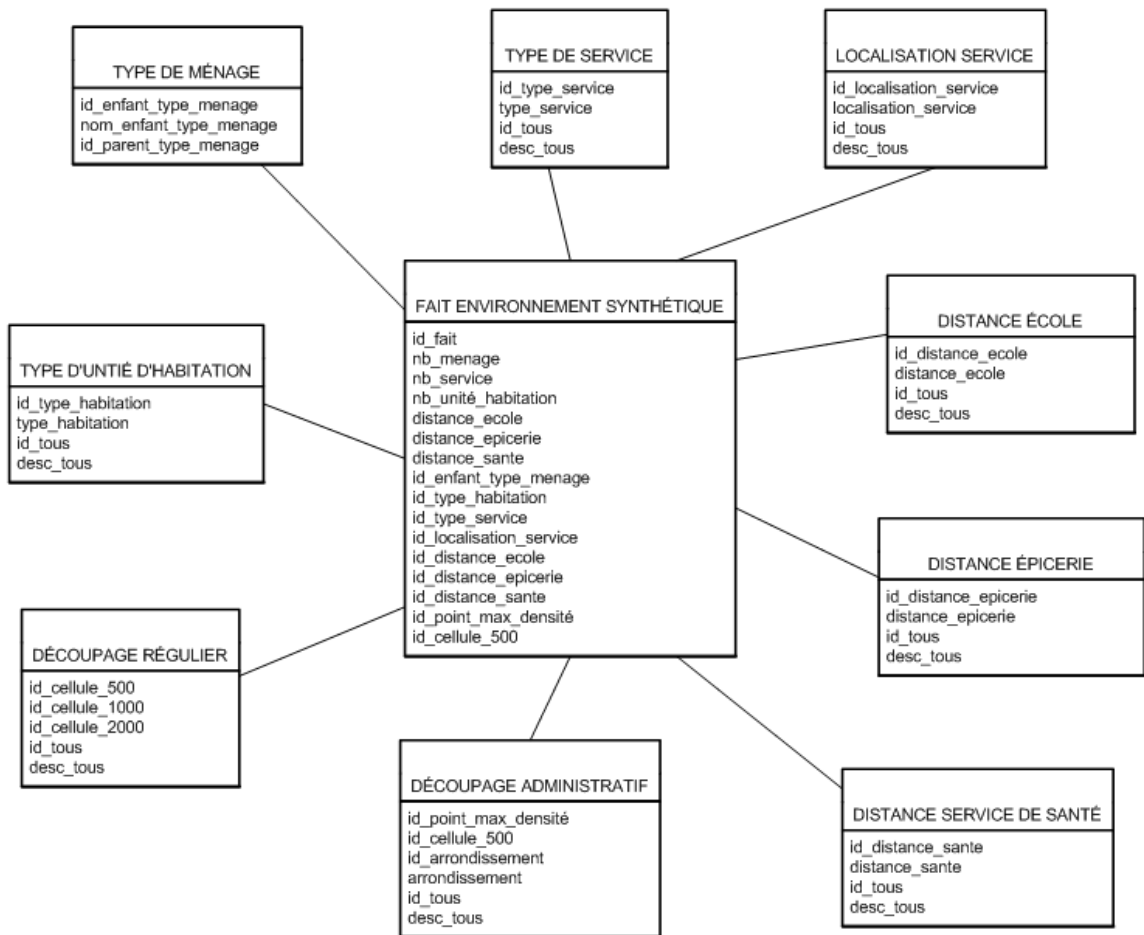


Figure A-1 Schéma étoile du cube Muscamags

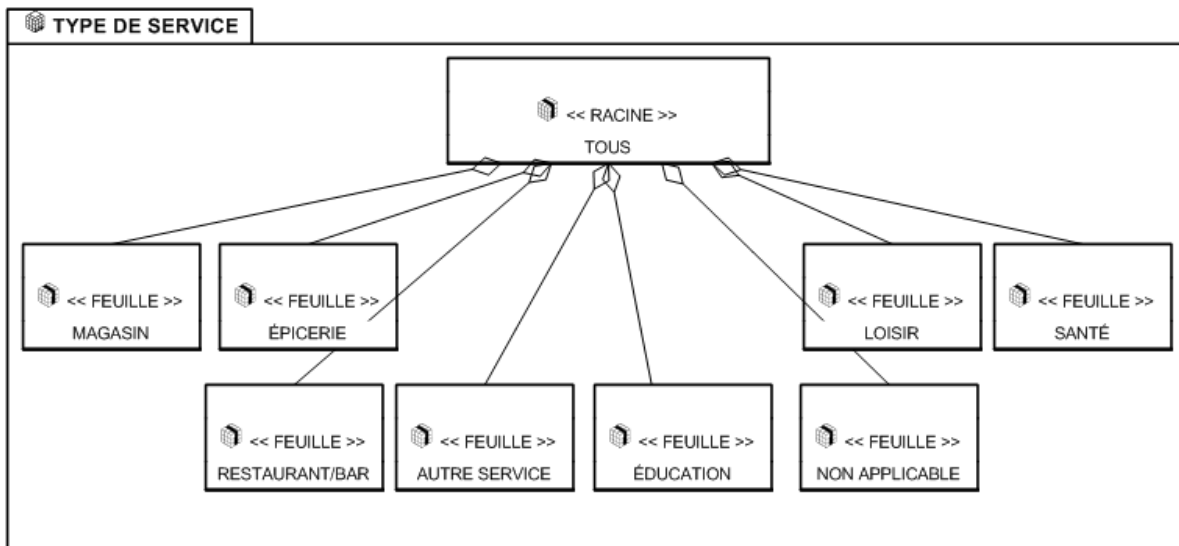


Figure A-2 Dimension type de service

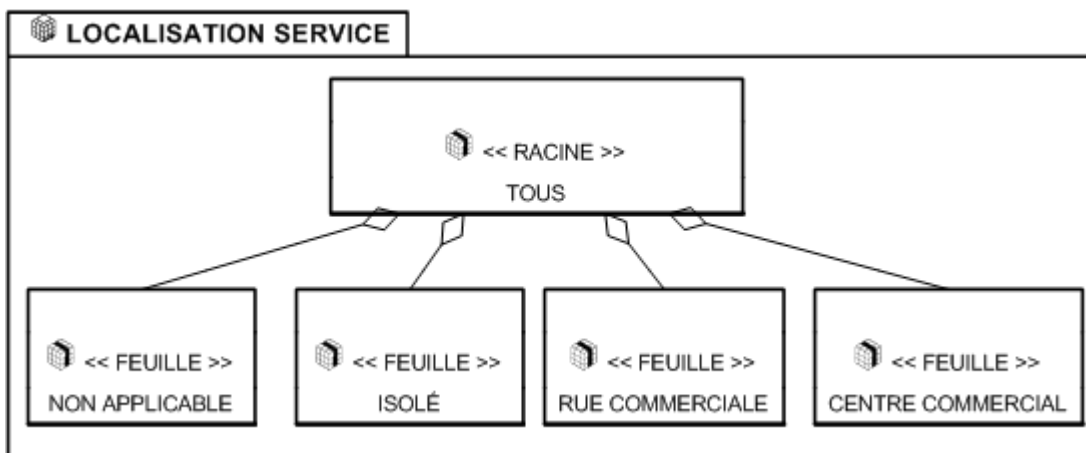


Figure A-3 Dimension localisation service

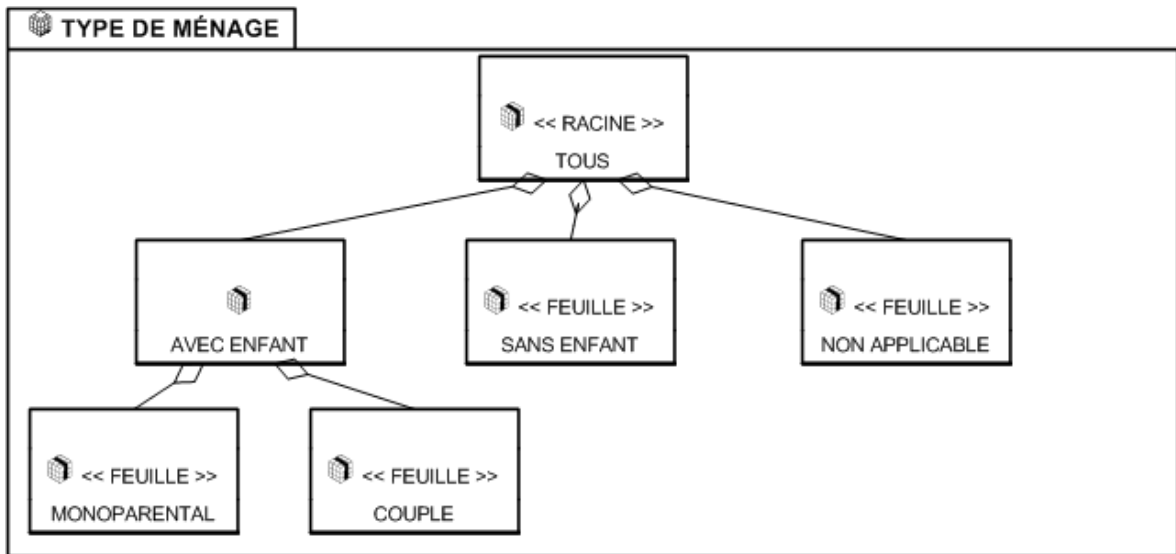


Figure A-4 Dimension type de ménage

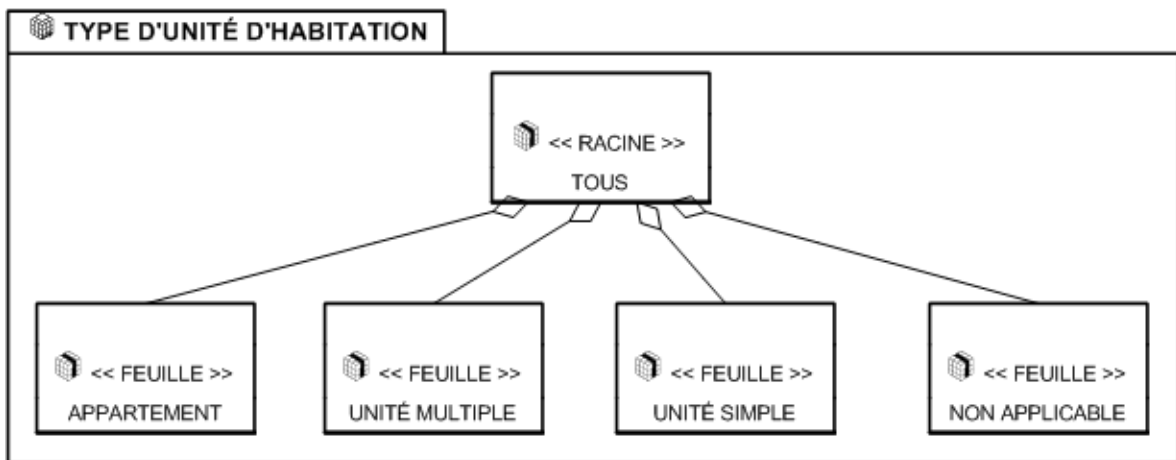


Figure A-5 Dimension type d'unité d'habitation

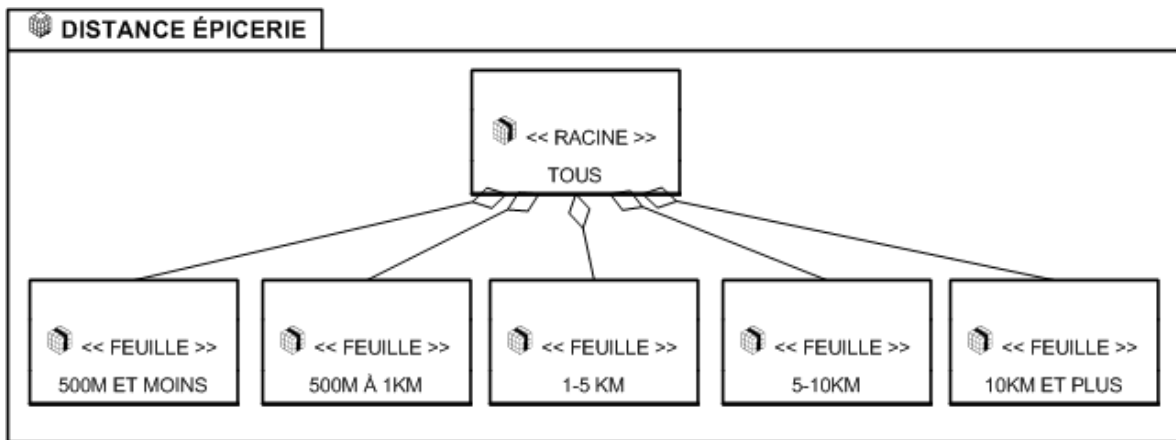


Figure A-6 Dimension distance épicerie

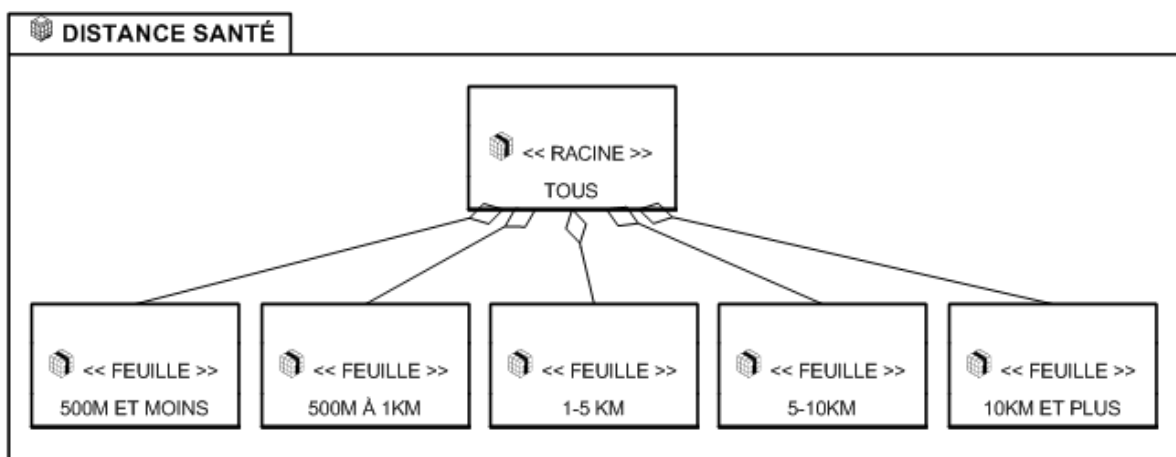


Figure A-7 Dimension distance santé

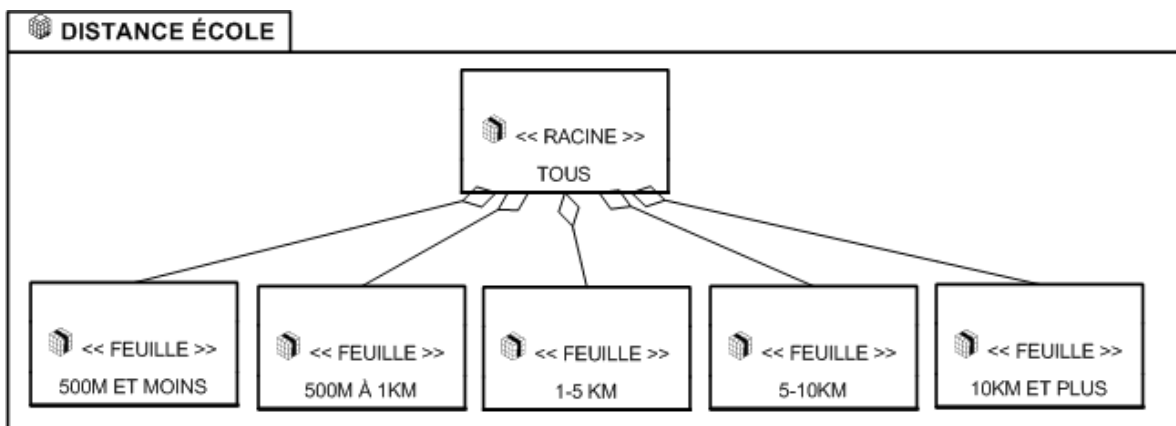


Figure A-8 Dimension distance école

Schémas expérimentation 2

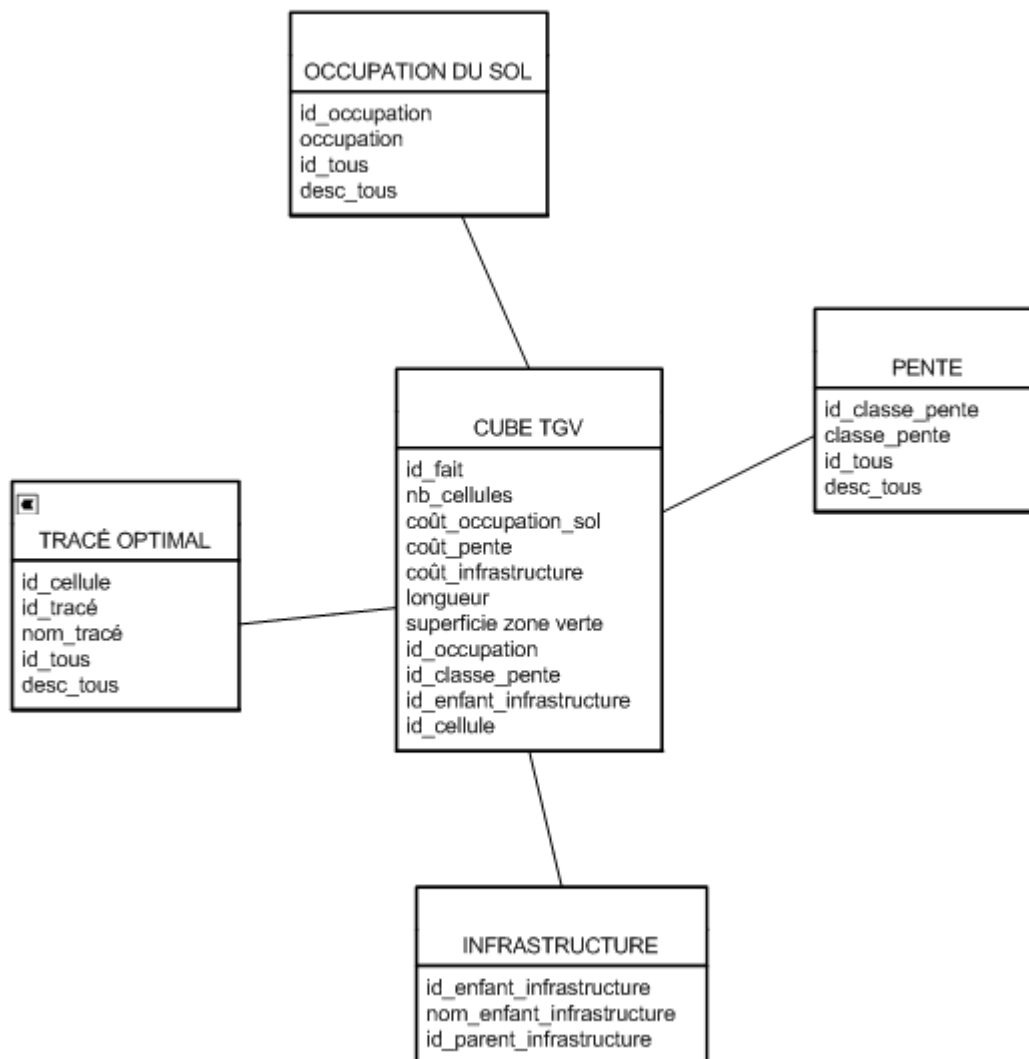


Figure A-9 Schéma étoile du cube TGV

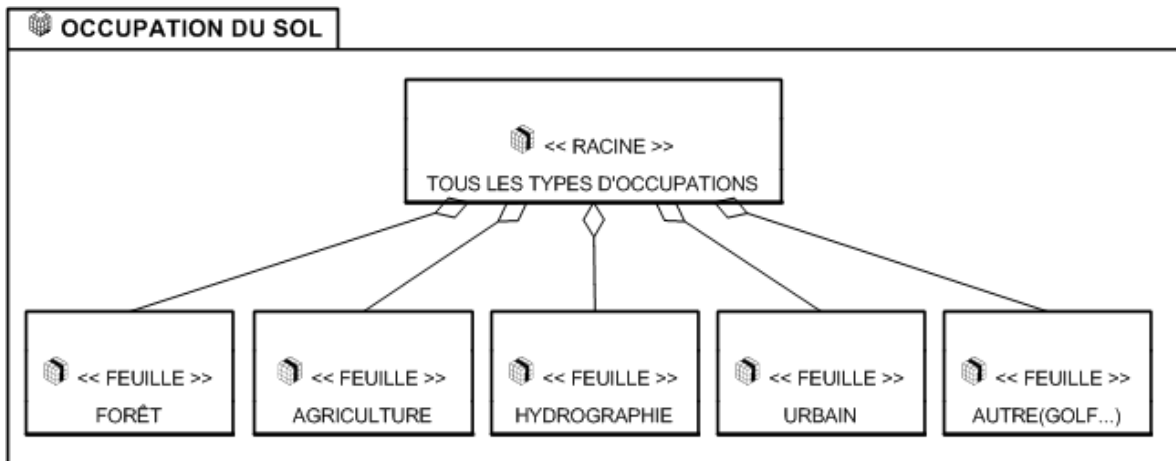


Figure A-10 Dimension occupation du sol

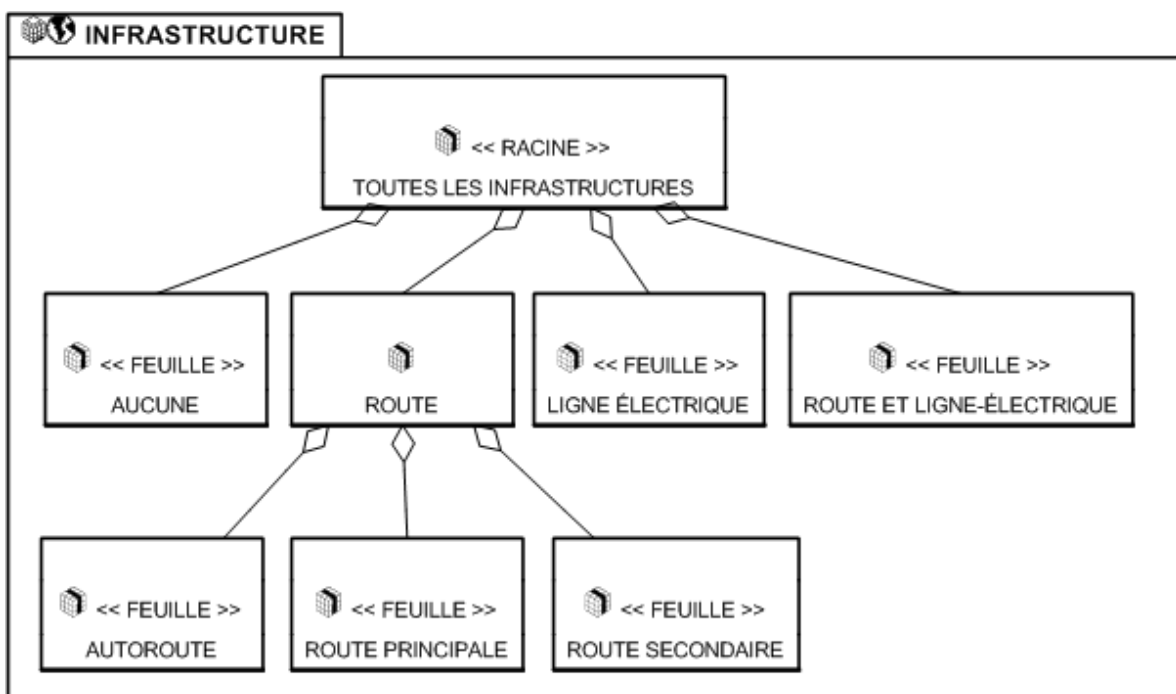


Figure A-11 Dimension infrastructure

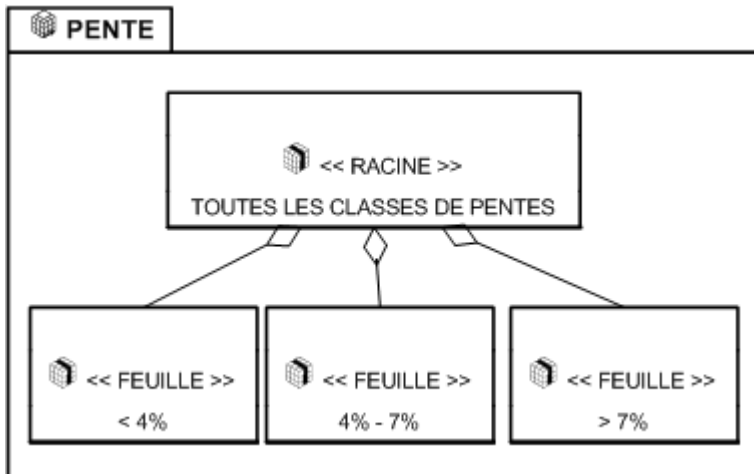


Figure A-12 Dimension pente

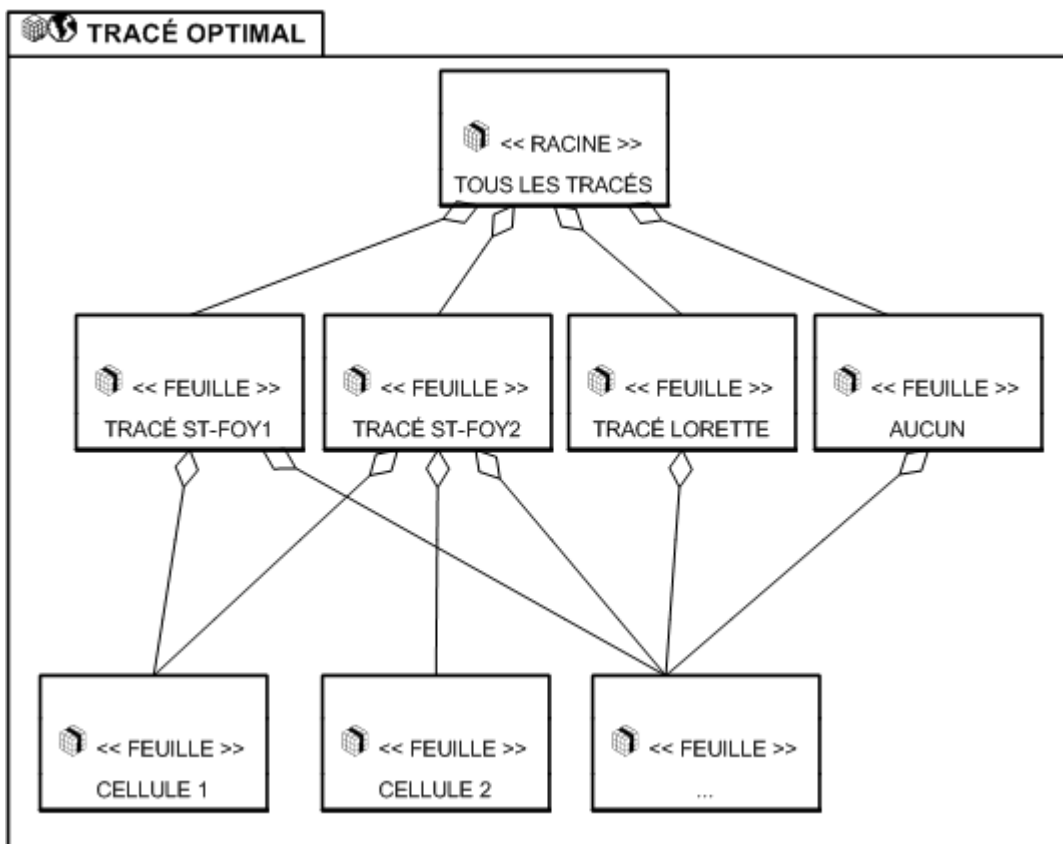


Figure A-13 Dimension tracé optimal

Schémas expérimentation formelle

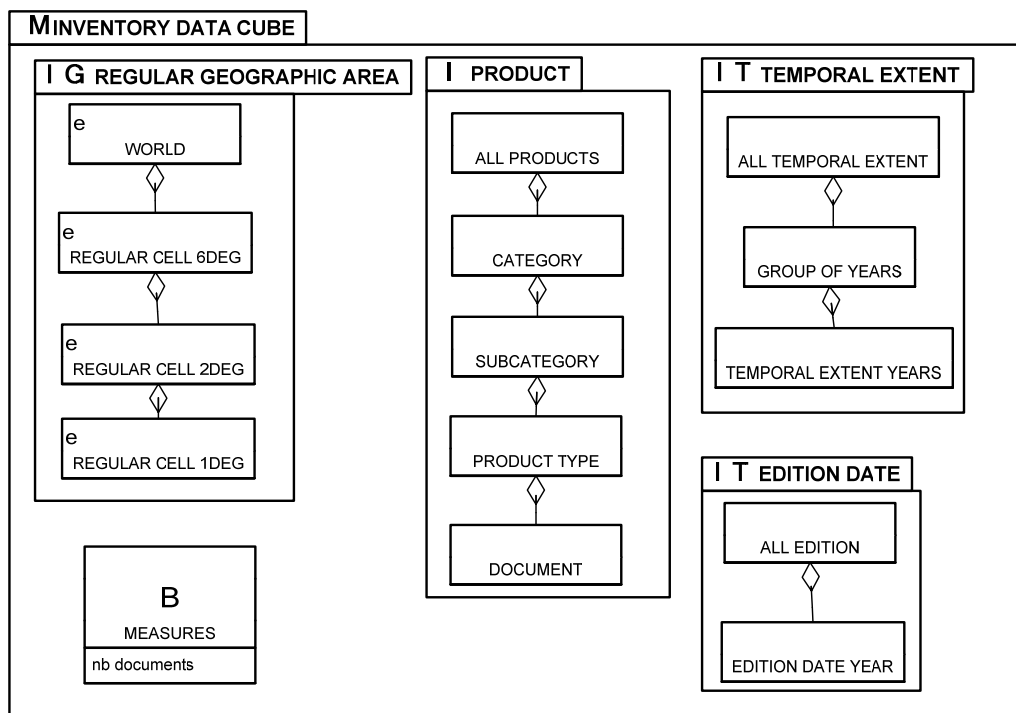


Figure A-14 Cube d'objets incluant une dimension spatiale géométrique matricielle (Source : [Proulx et *al.*, 2008])

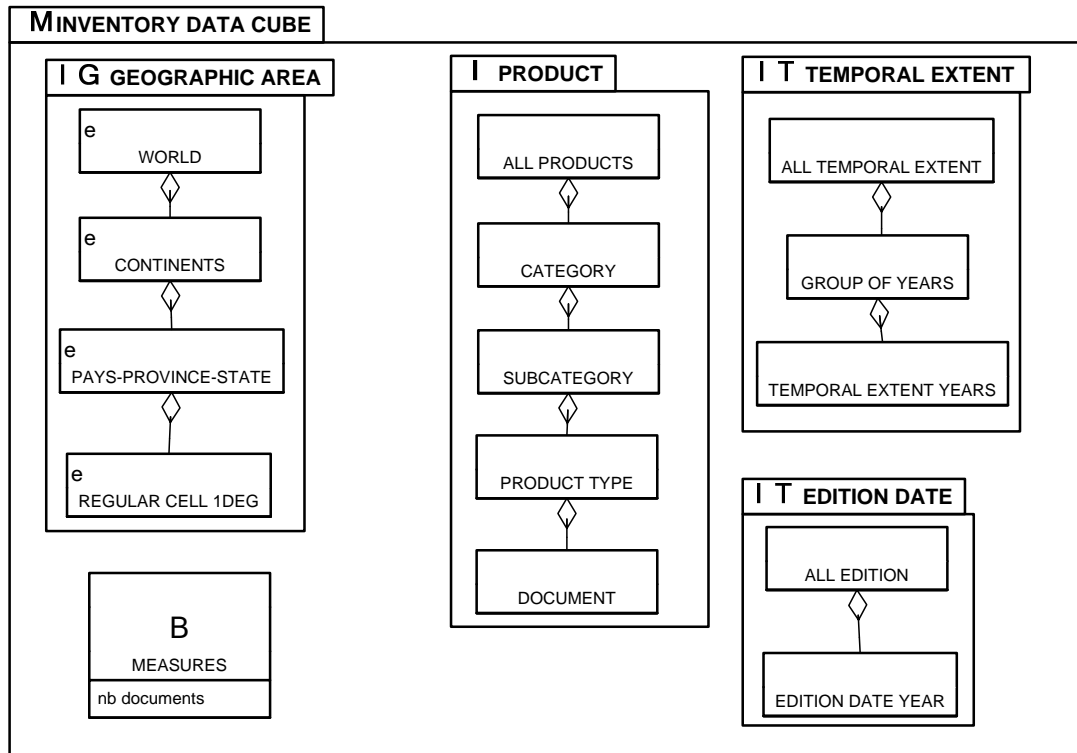


Figure A-15 Cube d'objets incluant une dimension spatiale géométrique hybride (Source : [Proulx et al., 2008])

ANNEXE B

Critères à respecter pour l'établissement des tracés de TGV (Expérimentation2)

CRITÈRES	INFORMATIONS COMPLÉMENTAIRES
1. Minimiser les coûts des terrains requis	<ul style="list-style-type: none"> • À cartographier à partir de l'image satellite. • Coûts du sol basés sur son occupation: <ul style="list-style-type: none"> - Agricole: 3 000\$/ha - Urbain: 60 000\$ /ha - Forêt: 12 000\$ / mille linéaire de corridor - Autres (ex. golf, aires désignées, marécages) : 1800\$/ha
2. Respecter l'emprise minimale requise pour tout le trajet du TGV	<ul style="list-style-type: none"> • Emprise: 166 pieds m.a.
3. Minimiser l'utilisation des bonnes terres agricoles (zones vertes)	<ul style="list-style-type: none"> • Carte régionale du zonage agricole – CPTAQ (Commission de la Protection du Territoire Agricole du Québec)
4. Minimiser les pentes	<ul style="list-style-type: none"> • Fichiers contenant les points cotés et les courbes de niveau • Intervalle à tenir compte : <ul style="list-style-type: none"> - < 4% aucun coût - entre 4% et 7% ajouter les coûts d'ajustement du relief, soit 30 000\$ par 100 mètres de longueur de terrain à ajuster - > 7% ajouter les coûts de construction d'un pont, soit 150 000\$ par 100 mètres de longueur
5. Minimiser le bruit et maximiser la sécurité	<ul style="list-style-type: none"> • Distance minimale de 500 mètres uniquement pour les villes de Donnacona, Neuville, Pont-Rouge, St-Augustin.
6. Éviter certaines infrastructures	<ul style="list-style-type: none"> • Certaines infrastructures ne peuvent pas être traversées et devront donc être contournées, telles les Lacs, les aéroports.
7. Ne pas passer dans la MRC "La Jacques-Cartier"	<p>Des pressions contre le TGV y ont été exercées</p> <ul style="list-style-type: none"> • Fichier des MRC de la région

CRITÈRES	INFORMATIONS COMPLÉMENTAIRES
8. Minimiser les coupures avec les infrastructures existantes	<ul style="list-style-type: none"> • Pour chaque type d'infrastructure, les coûts associés aux ajustements à effectuer sont : <ul style="list-style-type: none"> - Routes : 100 000\$ pour chaque tranche de 1 500 m² de contact (ici vous devez tenir compte du nombre de voies) - Cours d'eau : 150 000\$ pour chaque tranche de 1 500 m² de contact (ici vous devez tenir compte de la largeur du cours d'eau) - Lignes d'électricité : 80 000\$ pour chaque tranche de 30 m de contact

Tableau B-1 Critères à respecter (Sources : cours GMT-20853 Travaux pratiques en SIG et télédétection)

ANNEXE C

Opérateur de Map Algebra implanté dans Mondrian (Expérimentation 3)

```

/**UserDefined fonction qui permet de faire une fonction de Map Algebra
FocalRating
 * @author Rose
 * version 1.0 9 mars 2007
 */

import mondrian.olap.type.*;
import mondrian.spi.UserDefinedFunction;
import mondrian.olap.Evaluator;
import mondrian.olap.Syntax;
import java.lang.Object;

public class FocalRating implements UserDefinedFunction{

    public FocalRating(){
    }
    public String getName(){
        return "FocalRating";
    }

    public String getDescription(){
        return "Retourne le résultat de l'adjacence entre les éléments
passés arguments";
    }

    public Syntax getSyntax(){
        return Syntax.Function;
    }

    public Type getReturnType(Type[] parameterTypes){
        return new NumericType();
    }

    public Type[] getParameterTypes(){
        return new Type[]{new NumericType(), new NumericType(), new
NumericType(), new NumericType(),new NumericType(), new
NumericType(), new NumericType(), new NumericType(), new
NumericType(),new NumericType(), new NumericType()};
    }

```

```
}
public String[] getReservedWords() {
    return null;
}

public Object execute(Evaluator evaluator,
UserDefinedFunction.Argument[] arguments){
    final Object current = arguments[0].evaluateScalar(evaluator);
    final Object prev = arguments[1].evaluateScalar(evaluator);
    final Object next = arguments[2].evaluateScalar(evaluator);
    final Object voisin1 = arguments[3].evaluateScalar(evaluator);
    final Object voisin2 = arguments[4].evaluateScalar(evaluator);
    final Object voisin3 = arguments[5].evaluateScalar(evaluator);
    final Object voisin4 = arguments[6].evaluateScalar(evaluator);
    final Object voisin5 = arguments[7].evaluateScalar(evaluator);
    final Object voisin6 = arguments[8].evaluateScalar(evaluator);
    final Object typeSol1 = arguments[9].evaluateScalar(evaluator);
    final Object typeSol2 = arguments[10].evaluateScalar(evaluator);

    if(current.equals(typeSol2))
    {
        if(prev.equals(typeSol1)|| next.equals(typeSol1)||
voisin1.equals(typeSol1)||voisin2.equals(typeSol1)
||voisin3.equals(typeSol1)||voisin4.equals(typeSol1)||
voisin5.equals(typeSol1)||voisin6.equals(typeSol1))
        {
            return "1";
        }else{
            return "0";
        }
    }else{
        return "0";
    }
}
}
```

L'appel de cette fonction dans Mondrian s'effectue à l'aide d'une requête MDX telle que celle-ci :

```
WITH MEMBER [Measures].[Adjacence]
as IIf([Measures].[Type de sol]=0,0,FocalRating([Measures].[Type de sol],
([Measures].[Type de sol],[Colonne].PrevMember),
([Measures].[Type de sol],[Colonne].NextMember),
([Measures].[Type de sol],[Rangee].PrevMember),
([Measures].[Type de sol],[Rangee].NextMember),
([Measures].[Type de sol],[Colonne].PrevMember,[Rangee].PrevMember),
([Measures].[Type de sol],[Colonne].NextMember,[Rangee].PrevMember),
([Measures].[Type de sol],[Colonne].PrevMember,[Rangee].NextMember),
([Measures].[Type de sol],[Colonne].NextMember,[Rangee].NextMember),
2,100))
SELECT FILTER ([Colonne].members, [Colonne].CurrentMember NOT IN
{[Colonne].[1], [Colonne].LastChild}) on columns,
FILTER ([Rangee].members, [Rangee].CurrentMember NOT IN {[Rangee].[1],
[Rangee].LastChild}) on rows
FROM [Raster]
WHERE([Measures].[Adjacence])
```