

N° d'ordre 2007-ISAL-0105  
Année 2007

Thèse

# **Intégration de l'information géographique dans les entrepôts de données et l'analyse en ligne : de la modélisation à la visualisation**

Présentée devant  
L'Institut National des Sciences Appliquées de Lyon

Pour obtenir  
Le grade de docteur

Spécialité  
Informatique

Ecole doctorale  
Informatique et Information pour la Société

Par  
Sandro Bimonte

Soutenue le 18 Décembre 2007 devant la Commission d'examen

Jury MM.

M. Robert Laurini	Professeur à l'INSA-Lyon (Directeur de thèse)
M. Hervé Martin	Professeur à l'Université Joseph Fourier de Grenoble
M. Michel Schneider	Professeur à l'Université Blaise Pascal de Clermont-Ferrand (Rapporteur)
Mme Anne Tchounikine	Maître de Conférences à l'INSA-Lyon, (Codirectrice de thèse)
M. Esteban Zimányi	Professeur à l'Université Libre de Bruxelles, Belgique (Rapporteur)

Laboratoire de recherche  
Laboratoire d'InfoRmatique en Image et Systèmes d'information



Je tiens tout d'abord à remercier mes encadrant Anne Tchounikine et au Professeur Robert Laurini. Merci à Robert pour m'avoir permis d'entreprendre mes études doctorales à l'INSA de Lyon et merci à Anne pour son investissement humain et scientifique dans ce long travail de thèse. Elle a su me conduire dans ce travail avec capacité et conviction, en m'indiquant les bonnes pistes scientifiques, mais aussi en me laissant une importante liberté d'initiative. Merci alors pour votre suivi constant et pour l'amour de la recherche que vous m'avez transmis.

Je souhaite remercier les membres du jury. Merci au Professeur Michel Schneider et au Professeur Esteban Zimàny pour les conseils et les remarques qu'ils m'ont prodigués et qui m'ont permis d'améliorer mon travail. Je souhaite également remercier le Professeur Hervé Martin, tout d'abord pour m'avoir accueilli dans son équipe de travail cette année en tant qu'ATER, en me laissant le temps de finir la rédaction du manuscrit.

Merci aussi à toute l'équipe STEP, « Systèmes d'Information Spatio-Temporels & Entreposage », du LIRIS pour leurs conseils et remarques pertinentes. En particulier, je souhaite remercier Maryvonne Miquel, qui pendant ces trois années a fourni une contribution fondamentale à mon travail de thèse. Je tiens aussi à remercier mon co-équipier Pascal Wehrle, pour l'énorme aide qu'il m'a fourni dans la conception de mon prototype.

L'échange d'idées et le travail attentif de Sergio Di Martino et Filomena Ferrucci ont été fondamentaux pour mes recherches : merci. Un gros merci aussi à Beniamino Murgante pour m'avoir initié aux systèmes d'information géographiques.

La thèse est une aventure des plusieurs années pleines de joies, mais aussi de difficultés qu'il est impossible de surmonter tout seul. Je dois alors remercier énormément Julie pour son soutien moral infatigable et pour m'avoir appris la langue française. Je tiens aussi à remercier mes copains italiens Francesco, Rosario, Ruggero et Fabio pour avoir partagé des moments inoubliables ensemble ici en France.

Enfin, merci à mes parents d'avoir cru en moi.



# Résumé

L'informatique décisionnelle apporte des solutions nouvelles pour la modélisation, l'interrogation et la visualisation de données dans un objectif d'aide à la décision. Les modèles multidimensionnels ou modèles d'hypercube sont des modèles qui permettent de structurer les données pour l'analyse décisionnelle en explicitant la notion de dimension comme axe d'analyse (par exemple la dimension temps, la dimension produits, etc.), et de faits comme mesure ou indicateurs (par exemple le chiffre d'affaire ou le nombre d'articles vendus). L'algèbre OLAP (On Line Analytical Processing) formalise les opérateurs nécessaires à l'interrogation et à la navigation au sein des données multidimensionnelles. L'interrogation OLAP repose sur le parcours des hiérarchies de dimensions permettant de visualiser les mesures à différentes granularités de détail (par exemple les jours, les mois, les années, etc.). Des interfaces orientées navigation (tableau de bord, tableau multidimensionnels, graphes) complètent le panel des outils décisionnels.

L'intégration des données spatiales dans les systèmes OLAP est un enjeu majeur. En effet, l'information géographique est très fréquemment présente implicitement ou explicitement dans les données, mais généralement sous-employée dans le processus décisionnel. Le couplage de systèmes OLAP et de Systèmes d'Informations Géographiques au sein de systèmes OLAP Spatial (SOLAP) est une voie prometteuse.

La majorité des modèles et des outils SOLAP proposés en littérature réduisent l'information géographique à la seule composante spatiale, limitant ainsi les capacités d'analyse du paradigme spatio-multidimensionnel. Nous proposons un modèle formel (GeoCube) et une algèbre associée, qui reformule les concepts du SOLAP afin d'introduire les aspects sémantiques et spatiaux de l'information géographique dans l'analyse multidimensionnelle. Cela se traduit par une modélisation des mesures sous forme d'objets géographiques, dans une vision complètement symétrique entre mesures et dimensions. Ainsi une mesure peut participer à une hiérarchie de la même façon qu'un niveau de dimension. Nous proposons une algèbre qui fournit les opérateurs de forage et de coupe, un opérateur qui permet d'invertir mesure et dimension et des opérateurs de navigation au sein de la hiérarchie de mesures. Cette algèbre, grâce aux opérateurs qui modifient dynamiquement la structure de l'hypercube, permet de concilier analyse OLAP et analyse spatiale.

Nous avons réalisé un prototype web (GeWolap) conforme au modèle proposé. Pour décrire nos solutions, nous utilisons des données discrètes provenant d'études environnementales effectuées dans la lagune de Venise, notre travail de thèse s'inscrivant dans le cadre d'un projet international avec l'organisme CORILA (Consorzio per la Gestione del Centro di Coordinamento delle Attività di Ricerca inerenti il Sistema Lagunare di Venezia).

Enfin, nous nous intéressons à la visualisation des mesures géographiques dans un outil de navigation SOLAP. Nous proposons la conception d'un nouveau paradigme de visualisation et d'interaction pour l'analyse des mesures géographiques. Il combine les techniques OLAP classiques et une méthode de géovisualisation.

---

# Sommaire

---





## Table des matières

<b>INTRODUCTION</b> .....	<b>1</b>
<i>Contexte général de l'étude</i> .....	3
<i>Problématiques</i> .....	4
<i>Contributions des nos travaux de recherche</i> .....	5
<i>Organisation du mémoire</i> .....	6
<b>PARTIE 1 : ETAT DE L'ART</b> .....	<b>9</b>
CHAPITRE 1 :    ENTREPOTS DE DONNEES ET ANALYSE EN LIGNE.....	11
1.1 <i>Concepts principaux des entrepôts de données et de l'analyse en ligne</i> .....	11
1.1.1    Dimensions et hiérarchies.....	12
1.1.2    Faits et mesures .....	13
1.1.3    Hypercube.....	15
1.1.4    Requêtes multidimensionnelles.....	15
1.2 <i>Les modèles conceptuels pour les bases de données multidimensionnelles</i> .....	16
1.3 <i>Architecture des systèmes d'entrepôts de données</i> .....	18
1.3.1    Entrepôt de données.....	18
1.3.2    Serveur OLAP.....	18
1.3.2.1    Fonctionnalités d'un serveur OLAP .....	18
1.3.2.2    Implémentation physique d'un serveur OLAP .....	20
1.3.3    Client OLAP.....	21
1.4 <i>Conclusions</i> .....	23
CHAPITRE 2 :    L'INFORMATION GEOGRAPHIQUE ET LES SYSTEMES D'INFORMATION GEOGRAPHIQUE ...	25
2.1 <i>Les données géographiques</i> .....	25
2.1.1    La composante spatiale.....	25
2.1.2    La composante sémantique .....	26
2.1.3    La généralisation .....	26
2.2 <i>Fonctionnalités des Systèmes d'Information Géographique</i> .....	28
2.2.1    Les modèles matriciel et vectoriel, et les couches .....	28
2.2.2    Les bases de données spatiales.....	29
2.2.3    L'analyse spatiale.....	30
2.2.4    De la visualisation à la géovisualisation .....	32
2.2.4.1    La visualisation cartographique .....	32
2.2.4.2    La géovisualisation .....	34
2.3 <i>Conclusions</i> .....	36
CHAPITRE 3 :    L'OLAP SPATIAL .....	37
3.1 <i>De l'OLAP à l'OLAP Spatial</i> .....	37
3.1.1    Les systèmes d'aide à la décision spatiale.....	37
3.1.2    Les avantages de l'OLAP Spatial .....	38
3.2 <i>Concepts principaux de l'OLAP Spatial</i> .....	38
3.2.1    Dimension spatiale.....	39
3.2.1.1    Les approches existantes .....	39
3.2.1.2    Bilan.....	40
3.2.2    Mesure spatiale.....	40
3.2.2.1    La mesure spatiale vue comme une liste d'objets spatiaux.....	41
3.2.2.2    La mesure spatiale vue comme le résultat des opérateurs spatiaux .....	42
3.2.2.3    Bilan.....	43
3.2.3    Opérateurs de navigation spatio-multidimensionnelle .....	44
3.3 <i>Un panorama des modèles conceptuels pour les entrepôts de données spatiales</i> .....	44
3.3.1    Pourrabas .....	45
3.3.2    Malinowsky et Zimányi.....	45
3.3.3    Damiani et Spaccapietra .....	46
3.3.4    Jensen et al.....	47
3.3.5    Ahmed et Miquel .....	47
3.3.6    Bilan.....	48
3.4 <i>Les outils OLAP Spatial</i> .....	48

3.4.1	Typologies .....	48
3.4.1.1	SIG dominant .....	48
3.4.1.2	OLAP dominant .....	50
3.4.1.3	OLAP-SIG intégrée .....	50
3.4.2	Un panorama des outils SOLAP existants .....	52
3.4.2.1	Solutions SIG dominantes .....	52
3.4.2.2	Solutions OLAP dominantes .....	52
3.4.2.3	Solutions OLAP-SIG intégrées .....	53
3.4.3	Bilan .....	54
3.5	Conclusions .....	54
<b>PARTIE 2 : CONTRIBUTIONS.....</b>		<b>57</b>
CHAPITRE 4 :	DE L'OLAP SPATIAL A L'OLAP GEOGRAPHIQUE .....	59
4.1	Motivations .....	59
4.1.1	Modèle des données multidimensionnelles .....	60
4.1.1.1	Dimension spatiale .....	60
4.1.1.2	Mesure spatiale .....	61
4.1.2	Algèbre spatio-multidimensionnelle : flexibilité de l'analyse .....	62
4.2	Contexte .....	63
4.2.1	Cas d'étude : la pollution des eaux de la lagune de Venise .....	63
4.2.2	Notation graphique pour les modèles conceptuels applicatifs .....	64
4.3	Concepts principaux de l'OLAP Géographique .....	65
4.3.1	Dimension géographique .....	65
4.3.1.1	Hierarchie descriptive .....	66
4.3.1.2	Hierarchie spatiale .....	67
4.3.1.3	Hierarchie de généralisation .....	68
4.3.1.4	Problématiques .....	69
4.3.1.4.1	Modélisation .....	69
4.3.1.4.2	Visualisation .....	71
4.3.2	Mesure géographique .....	71
4.3.2.1	Définition et application .....	71
4.3.2.2	Problématiques .....	73
4.3.2.2.1	Modélisation .....	73
4.3.2.2.2	Visualisation .....	75
4.3.3	Opérateurs d'analyse spatio-multidimensionnelle .....	76
4.3.3.1	Modification dynamique de la structure de l'hypercube .....	77
4.3.3.2	Permutation entre mesure et dimension .....	77
4.3.3.3	Navigation dans la hiérarchie de la mesure .....	78
4.3.3.4	Problématiques .....	79
4.4	Conclusion .....	79
CHAPITRE 5 :	GEOCUBE, UN MODELE MULTIDIMENSIONNEL OLAP GEOGRAPHIQUE .....	81
5.1	Les données multidimensionnelles .....	81
5.1.1	Entité .....	81
5.1.2	Hierarchie .....	84
5.1.2.1	Définitions .....	85
5.1.2.2	Représentation graphique .....	87
5.1.3	Cube de Base .....	89
5.2	Requêtes multidimensionnelles .....	92
5.2.1	Mode d'Agrégation .....	92
5.2.2	Vue .....	95
5.3	Algèbre .....	100
5.3.1	Opérateurs de forage et de coupe .....	101
5.3.1.1	Roll-Up .....	101
5.3.1.2	Slice .....	104
5.3.1.3	Dice .....	109
5.3.2	Opérateurs de navigation dans la hiérarchie de la mesure .....	112
5.3.2.1	Classify .....	114
5.3.2.2	Specialize .....	118
5.3.3	Opérateur de permutation mesure et dimension : Permute .....	121
5.3.4	Opérateur de modification dynamique de l'hypercube : OLAP-Buffer .....	122
5.4	Requêtes multidimensionnelles complexes .....	128

5.4.1	Reformulation du processus d'agrégation par le changement de la granularité de la mesure .....	128
5.4.2	Roll-up sur des cousins.....	130
5.5	<i>Bilan et comparaison</i> .....	132
5.5.1	Caractéristiques principales de GeoCube.....	132
5.5.2	Comparaison.....	134
5.5.3	Limites .....	135
5.6	<i>Conclusions</i> .....	135
CHAPITRE 6 : GEWOLAP, UNE SOLUTION WEB OLAP GEOGRAPHIQUE.....		137
6.1	<i>Architecture</i> .....	137
6.1.1	Entrepôt de données spatiales .....	139
6.1.2	Serveur OLAP.....	139
6.1.3	Client OLAP.....	142
6.2	<i>L'interface visuelle de GeWolap</i> .....	143
6.2.1	Le Data Panel .....	144
6.2.1.1	Toolbar SIG.....	144
6.2.1.2	Toolbar multidimensionnelle non-spatiale.....	145
6.2.1.3	Toolbar multidimensionnelle spatiale .....	145
6.2.2	Le Control Panel .....	146
6.3	<i>Dimension Géographique et opérateurs multidimensionnels</i> .....	147
6.3.1	Opérateurs multidimensionnels non-spatiaux .....	150
6.3.2	Opérateurs multidimensionnels spatiaux.....	153
6.3.2.1	Opérateurs de forage.....	153
6.3.2.2	Opérateurs de coupe .....	154
6.3.2.3	Opérateurs de changement dynamique de l'hypercube.....	156
6.4	<i>Mesures complexes et géographiques, et opérateurs multidimensionnels</i> .....	159
6.4.1	Modélisation logique .....	160
6.4.2	Visualisation et Navigation .....	164
6.4.2.1	Mesures géographiques .....	164
6.4.2.2	Mesures complexes.....	166
6.5	<i>Opérateurs SIG</i> .....	169
6.6	<i>Bilan et comparaison</i> .....	169
6.6.1	Caractéristiques principales.....	169
6.6.2	Comparaison et limites .....	170
6.7	<i>Conclusions</i> .....	171
CHAPITRE 7 : GEOLAPIVOT TABLE, UNE TECHNIQUE DE VISUALISATION ET D'INTERACTION POUR L'ANALYSE DES MESURES GEOGRAPHIQUES .....		173
7.1	<i>Pivot Table et Space-Time Cube</i> .....	173
7.2	<i>GeOlapivot Table</i> .....	175
7.2.1	Le concept .....	175
7.2.2	Un cas d'étude.....	176
7.3	<i>Caractéristiques principales et limitations</i> .....	181
7.4	<i>Conclusions</i> .....	181
<b>PARTIE 3 : CONCLUSIONS .....</b>		<b>183</b>
CHAPITRE 8 : BILAN ET PERSPECTIVES .....		185
8.1	<i>Bilan des travaux</i> .....	185
8.1.1	Le paradigme OLAP Géographique .....	185
8.1.2	Le modèle GeoCube .....	186
8.1.3	Le prototype GeWolap .....	186
8.1.4	GeOlapivot Table.....	187
8.2	<i>Perspectives</i> .....	187
8.2.1	Additivité des mesures géographiques .....	187
8.2.2	Hierarchies de généralisation .....	188
8.2.3	Composante temporelle de l'information géographique .....	188
8.2.4	Mesures continues.....	189
8.2.5	Géovisualisation et OLAP Géographique .....	189
8.2.6	L'OLAP Géographique Mobile.....	190
REFERENCES.....		191

## Liste des tableaux

TABLEAU 1.1. DIFFERENCES PRINCIPALES ENTRE LES SYSTEMES OLPT ET OLAP (KELLY, 1997).	12
TABLEAU 2.1. VARIABLES VISUELLES (BERTIN ET BONIN, 1992).	34
TABLEAU 5.1. INSTANCE DU CUBE DE BASE $BC_{CORILA}$ .	92
TABLEAU 5.2. INSTANCE DE $V_{CORILA-UNIT}$ .	98
TABLEAU 5.3. INSTANCE DE $V_{CORILA-UNIT-MONTH}$ .	100
TABLEAU 5.4. INSTANCE DE $V_{CORILA-UNIT-1000}$ .	107
TABLEAU 5.5. INSTANCE DE $V_{CORILA-UNIT-MONTH-1500}$ .	108
TABLEAU 5.6. INSTANCE DE $BC_{CORILA1500}$ .	109
TABLEAU 5.7. INSTANCE DE $V_{CORILA-UNIT/TIME}$ .	112
TABLEAU 5.8. INSTANCE DE $V_{CORILA-ZONE}$ .	117
TABLEAU 5.9. INSTANCE DE $V_{CORILA-ZONE-1000}$ .	118
TABLEAU 5.10. INSTANCE DE $V_{CORILA-ZONE-1000S}$ .	121
TABLEAU 5.11. INSTANCE DE $V_{RATE}$ .	127
TABLEAU 5.12. INSTANCES DE $V_{RATEB}$ .	127
TABLEAU 5.13. INSTANCE DE $V_{CORILA-MONTH}$ .	130
TABLEAU 5.14. INSTANCE DE $V_{CORILA-MONTH-ZONE}$ .	130
TABLEAU 5.15. SLICE SUR MAZZORBO ET CANAL FONDELLO.	131
TABLEAU 5.16. AGREGATION ENTRE COUSINS.	132
TABLE 7.1. EXEMPLE DE LA TABLE DE FAITS DE L'APPLICATION MULTIDIMENSIONNELLE DE FIGURE 7.4.	177

## Liste des figures

FIGURE 1.1. HIERARCHIE DES PRODUITS A) SCHEMA B) INSTANCE.	13
FIGURE 1.2. APPLICATION MULTIDIMENSIONNELLE A) SCHEMA B) HYPERCUBE.	14
FIGURE 1.3. ARCHITECTURE A 3-NIVEAUX D'UN SYSTEME D'ENTREPOT DE DONNEES.	18
FIGURE 1.4. A) HIERARCHIE AVEC LES NIVEAUX « A » ET « B » B) HIERARCHIE AVEC LES NIVEAUX « 1 » ET « 2 » C) TREILLIS DE CUBOÏDES.	19
FIGURE 1.5. A) SCHEMA EN ETOILE B) SCHEMA EN FLOCON C) SCHEMA EN CONSTELLATION.	21
FIGURE 1.6. CLIENTS OLAP A) MICROSTRATEGY B) COGNOS C) ORACLE.	22
FIGURE 1.7. VISUALISATION DES DONNEES DE L'APPLICATION MULTIDIMENSIONNELLE DES VENTES DE FIGURE 1.2 AVEC LA TABLE DE PIVOT DE JPivot.	23
FIGURE 2.1. LA COMPOSANTE SPATIALE DES DEPARTEMENTS DE FRANCE.	26
FIGURE 2.2. EXEMPLIS DE RELATIONS D'IDENTITE ET DE GROUPE (NEUN ET STEINIGER, 2005).	27
FIGURE 2.3. ORGANISATION DES INFORMATIONS GEOGRAPHIQUES EN COUCHES.	29
FIGURE 2.4. ORACLE SPATIAL A) CREATION D'UNE TABLE SPATIALE B) REQUETE SQL QUI UTILISE EN PREDICAT SPATIAL C) INSERTION D'UN OBJET GEOGRAPHIQUE DANS LA TABLE D) REPRESENTATION TABULAIRE ET CARTOGRAPHIQUE.	30
FIGURE 2.5. PROCESSUS D'ANALYSE SPATIALE (MITCHELL, 2005).	32
FIGURE 2.6. MAP-USE-CUBE (MACEachREN ET KRAAK, 1997).	35
FIGURE 2.7. A) SPACE-TIME CUBE (KRAAK ET KOUSSOULAKOU, 2005) B) MATRICE MULTIFORME BIVARIEE (MULTIFORM BIVARIATE MATRIX) (MACEachREN, ET AL., 2004).	36
FIGURE 3.1. DIMENSION SPATIALE GEOMETRIQUE A) SCHEMA B) REPRESENTATION CARTOGRAPHIQUE DES MEMBRES.	39
FIGURE 3.2. SCHEMA DE L'ENTREPOT « METEO » (STEFANOVIC ET AL., 2000).	41
FIGURE 3.3. APPLICATION SPATIO-MULTIDIMENSIONNELLE AVEC DE DIMENSIONS ALPHANUMERIQUES ET UNE MESURE SPATIALE (MALINOWSKI ET ZIMANYI, 2004).	42
FIGURE 3.4. SCHEMA CONCEPTUEL D'UN ENTREPOT DE DONNEES AVEC UNE MESURE RESULTAT D'UNE OPERATION SPATIALE (MALINOWSKI ET ZIMANYI, 2004).	43
FIGURE 3.5. REPRESENTATION GRAPHIQUE DU MODELE POUR ENTREPOTS DE DONNEES SPATIALES PRESENTE DANS (MALINOWSKI ET ZIMANYI, 2004) A) FAIT ET MESURES B) DIMENSION C) CARDINALITES D) NIVEAUX DE DIMENSION E) ICONES SPATIALES.	46

FIGURE 3.6. INTERFACES DES SYSTEMES SOLAP A) COMMONGIS (HERNANDEZ ET AL., 2005) B) POLARIS (STOLTE ET AL., 2002) C) POSTGEO LAP (COLONESE ET AL., 2005) D) JMAP (JMAP, 2007) E) SOVAT (SCOTCH ET PARMANTO, 2006).	49
FIGURE 4.1. CARTE DE LA LAGUNE DE VENISE.	64
FIGURE 4.2. EXTENSION DE LA NOTATION GRAPHIQUE POUR LES ENTREPOTS DE DONNEES SPATIALES PRESENTEE DANS (MALINOWSKY, 2006) AVEC UNE MESURE GEOGRAPHIQUE ET DES ATTRIBUTS DERIVES.	65
FIGURE 4.3. HIERARCHIE DESCRIPTIVE A) SCHEMA B) INSTANCE.	66
FIGURE 4.4. APPLICATION MULTIDIMENSIONNELLE AVEC UNE DIMENSION GEOGRAPHIQUE DECRITE PAR UNE HIERARCHIE DESCRIPTIVE.	66
FIGURE 4.5. HIERARCHIE SPATIALE A) SCHEMA B) INSTANCE.	67
FIGURE 4.6. APPLICATION MULTIDIMENSIONNELLE AVEC UNE DIMENSION GEOGRAPHIQUE DECRITE PAR UNE HIERARCHIE SPATIALE.	68
FIGURE 4.7. APPLICATION MULTIDIMENSIONNELLE AVEC UNE DIMENSION GEOGRAPHIQUE DECRITE PAR UNE HIERARCHIE DE GENERALISATION.	68
FIGURE 4.8. HIERARCHIE DE GENERALISATION A) INSTANCE B) REPRESENTATION CARTOGRAPHIQUE.	69
FIGURE 4.9. HIERARCHIE DE GENERALISATION A) SCHEMA B) INSTANCES ET RELATION MULTI-ASSOCIATION.	70
FIGURE 4.10. MESURE GEOGRAPHIQUE « UNIT ».	72
FIGURE 4.11. MESURE GEOGRAPHIQUE « UNIT » A L'ECHELLE 1 : 500.	73
FIGURE 4.12. OVERLAY UTILISANT L'OPERATEUR LOGIQUE AND.	77
FIGURE 4.13. PERMUTATION MESURE ET DIMENSION : DE LA MESURE GEOGRAPHIQUE A LA DIMENSION GEOGRAPHIQUE.	78
FIGURE 5.1. REPRESENTATION GRAPHIQUE A) SCHEMA D'ENTITE B) ORDRE PARTIEL DES SCHEMAS D'ENTITE NIVEAUX $\mathbb{F}_H$ C) ORDRE PARTIEL DES INSTANCES DES SCHEMAS D'ENTITE NIVEAUX $\uparrow_H$ .	87
FIGURE 5.2. HIERARCHIE DE POLLUANTS A) SCHEMA B) INSTANCE.	88
FIGURE 5.3. HIERARCHIE SPATIALE DE LA LAGUNE A) SCHEMA B) INSTANCE.	88
FIGURE 5.4. REPRESENTATION DE LA FONCTION BOOLEENNE D'UN CUBE DE BASE AVEC TROIS SCHEMAS DE HIERARCHIE.	90
FIGURE 5.6. MODE D'AGREGATION POUR L'AGREGATION DES OBJETS GEOGRAPHIQUES REPRESENTANTS LES UNITES DE LA LAGUNE.	94
FIGURE 5.7. REPRESENTATION GRAPHIQUE DU MODE D'AGREGATION.	95
FIGURE 5.8. REPRESENTATION GRAPHIQUE DU SCHEMA DE VUE A) GRAPHE D'AGREGATION B) GRAPHE DE NAVIGATION C) NIVEAUX DE LA VUE D) MESURE DETAILLEE OU MESURE AGREGEE.	97
FIGURE 5.9. REPRESENTATION GRAPHIQUE DU SCHEMA DE VUE $V_{CORILA-UNIT}$ (VUE DE BASE).	98
FIGURE 5.10. REPRESENTATION GRAPHIQUE DU SCHEMA DE VUE $V_{CORILA-POLLUTANT}$ (VUE DE BASE).	99
FIGURE 5.11. REPRESENTATION GRAPHIQUE DU SCHEMA DE VUE $V_{CORILA-UNIT-MONTH}$ .	100
FIGURE 5.12. REPRESENTATION GRAPHIQUE DE $V_{CORILA-UNIT/TIME}$ .	111
FIGURE 5.13. A) SCHEMA DE HIERARCHIE $H_A$ ET INSTANCE DE $H_A$ B) HIERARCHIE REDUITE (SCHEMA ET INSTANCE) DE $H_A$ SUR $S_B$ .	114
FIGURE 5.14. HIERARCHIE REDUITE DE $H_{LAGOON\_SPATIAL}$ SUR $S_{ZONE}$ A) SCHEMA B) INSTANCE.	114
FIGURE 5.15. REPRESENTATION GRAPHIQUE DE $V_{CORILA-ZONE}$ .	117
FIGURE 5.16. EXEMPLES D'APPLICATION DE L'OPERATEUR BUFFER.	123
FIGURE 5.17. REPRESENTATION CARTOGRAPHIQUE A) $S_{ZONE}$ B) $S_{ZONEB}$ .	125
FIGURE 5.18. REPRESENTATION GRAPHIQUE A) $H_{LAGOON\_SPATIAL\_ZONE}$ B) $H_{LAGOON\_SPATIAL\_ZONEB}$ .	126
FIGURE 5.19. AGREGATION SUR LES COUSINS : MAZZORBO ET CANAL FONDELLO.	131
FIGURE 6.1. ARCHITECTURE DE GEWOLAP.	138
FIGURE 6.2. FICHER XML REPRESENTANT LE SCHEMA MONDRIAN D'UNE APPLICATION MULTIDIMENSIONNELLE.	141
FIGURE 6.3. REQUETE MULTIDIMENSIONNELLE A) MDX B) TRADUCTION EN SQL.	142
FIGURE 6.4. DEFINITION D'UN AGGREGATE TABLE.	142
FIGURE 6.5. INTERFACE VISUELLE DE GEWOLAP : CONTROL PANEL ET DATA PANEL.	144
FIGURE 6.6. TOOLBAR SIG.	145
FIGURE 6.7. TOOLBAR MULTIDIMENSIONNELLE NON-SPATIALE.	145
FIGURE 6.8. TOOLBAR MULTIDIMENSIONNELLE SPATIALE.	146
FIGURE 6.9. CONTROL PANEL A) NAVIGATION B) DISPLAY EXPORT C) ANALYSIS.	147
FIGURE 6.10. SCHEMA CONCEPTUEL D'UNE APPLICATION AVEC UNE HIERARCHIE SPATIALE.	147
FIGURE 6.11. SCHEMA MONDRIAN DE L'APPLICATION DE FIGURE 6.10.	149

FIGURE 6.12. SELECTION DES MEMBRES DE LA DIMENSION DES POLLUANTS ET DE LA DIMENSION GEOGRAPHIQUE A TRAVERS LE CUBE NAVIGATOR. ....	150
FIGURE 6.13. REQUETE MDX GENEREE PAR JPivot A TRAVERS L'INTERACTION AVEC LA TABLE DE PIVOT...	150
FIGURE 6.14. VISUALISATION DE LA REQUETE MULTIDIMENSIONNELLE DE FIGURE 6.13. ....	151
FIGURE 6.15. RELATIONS TOPOLOGIQUES ET SYNCHRONISATION DE LA COMPOSANTE CARTOGRAPHIQUE ET DE LA TABLE DE PIVOT. ....	152
FIGURE 6.16. DRILL-DOWN SUR UNE DIMENSION CLASSIQUE ET SYNCHRONISATION DE LA COMPOSANTE CARTOGRAPHIQUE ET DE LA TABLE DE PIVOT. ....	153
FIGURE 6.17. GEWOLAP DRILL-DOWN POSITION. ....	154
FIGURE 6.18. GEWOLAP SLICE PREDICATE : 1) DONNEES INITIALES : TOUTES LES REGION DE LA LAGUNE 2) RESULTAT : REGIONS DE LA LAGUNE AVEC UNE SURFACE SUPERIEURE A 1000 KM <sup>2</sup> . ....	155
FIGURE 6.19. REQUETE SQL REPRESENTANT LE PREDICAT SPATIAL D'UNE OPERATION DE COUPE.....	156
FIGURE 6.20. GEWOLAP BUFFER 1) DONNEES INITIALES 2) RESULTAT. ....	157
FIGURE 6.21. 1) DONNEES INITIALES 2) RESULTATS DU GEWOLAP OVERLAY. ....	158
FIGURE 6.22. SCHEMA ISSU DE L'OPERATION DE GEWOLAP OVERLAY.....	159
FIGURE 6.23. APPLICATION MULTIDIMENSIONNELLE AVEC LA MESURE GEOGRAPHIQUE « UNIT ». ....	159
FIGURE 6.24. APPLICATION MULTIDIMENSIONNELLE AVEC LA MESURE COMPLEXE « POLLUTANT » ET LA DIMENSION GEOGRAPHIQUE « LAGOON ». ....	160
FIGURE 6.25. SQUELETTE DU SCHEMA UTILISE POUR LA MODELISATION DES APPLICATIONS DES FIGURES 6.23 ET 6.24.....	161
FIGURE 6.26. BALISE MEASURE A) MESURE GEOGRAPHIQUE B) MESURE COMPLEXE. ....	162
FIGURE 6.27. PROCESSUS D'AGREGATION DES MESURES COMPLEXES ET GEOGRAPHIQUES. ....	162
FIGURE 6.28. AGGREGATE TABLES POUR LA GESTION DE MESURES MULTIVALUEES A) « UNIT » B) « POLLUTANT ». ....	164
FIGURE 6.29. MESURE GEOGRAPHIQUE « UNIT ». ....	165
FIGURE 6.30. AGREGATION DES MESURES GEOGRAPHIQUES « UNIT ». ....	166
FIGURE 6.31. TABLE DE PIVOT POUR LA MESURE COMPLEXE « POLLUTANT ». ....	167
FIGURE 6.32. CARTE CHOROPLETHE POUR LA MESURE COMPLEXE « POLLUTANT » ASSOCIEE AU JOUR 19/9/05 ET A LA VALEUR DE POLLUTION 30-34 MG/L.....	168
FIGURE 6.33. CARTE CHOROPLETHE POUR LA MESURE COMPLEXE « POLLUTANT » POUR LE JOUR 26/9/05 ET LA VALEUR DE POLLUTION 30-34 MG/L.....	168
FIGURE 7.1 . EXTENSION DE LA TABLE DE PIVOT AVEC DES AFFICHAGES GRAPHIQUES ET DES VARIABLES VISUELLES DANS TABLEAU SOFTWARE. ....	174
FIGURE 7.2. SPACE-TIME CUBE A) ETUDE DES TRAJECTOIRES (KRAAK, 2003) B) ETUDE DES TREMBLEMENTS DE TERRE (GATALSKY ET AL., 2004).....	175
FIGURE 7.3. EXEMPLE DE GEOLAPIVOT TABLE. ....	176
FIGURE 7.4. APPLICATION MULTIDIMENSIONNELLE AVEC UNE MESURE GEOGRAPHIQUE « DISTRICT » ET UNE DIMENSION GEOGRAPHIQUE « LOCATION ». ....	177
FIGURE 7.5. MAQUETTE DE L'INTERFACE QUI UTILISE LA GEOLAPIVOT TABLE. ....	179
FIGURE 7.6. DRILL DOWN SUR LA DIMENSION TEMPORELLE ET LA DIMENSION GEOGRAPHIQUE.....	180

---

# Introduction

---





## Contexte général de l'étude

Les entrepôts de données associés à des outils d'analyse On Line Analytical Processing (OLAP), représentent une solution effective pour l'informatique décisionnelle (Immon, 1992). Les données dans les hypercubes sont organisées en axes d'analyses appelés dimensions. Les sujets d'analyse, appelés « faits » sont caractérisés par des mesures, qui sont pré-calculées à l'aide de fonctions d'agrégation selon les différentes granularités définies par le schéma hiérarchique de chaque dimension. Dans le cas classique, une mesure est une valeur numérique qui décrit quantitativement le fait. Ainsi une analyse multidimensionnelle portant sur un fait « ventes » d'un ensemble de magasins pourra être réalisée en définissant comme mesures « le volume de la vente » et « le montant de la vente ». Le processus d'analyse est conduit par la navigation dans le cube multidimensionnel. Les opérateurs OLAP (roll-up, drill-down, slice, etc.) permettent de visualiser les mesures pour des ensembles de membres à des niveaux de granularité sélectionnés par l'utilisateur. Les opérations de forage (roll-up, drill-down) sont fondées sur des fonctions d'agrégation appliquées aux mesures, par exemple la somme appliquée au volume de produits vendus. Des interfaces orientées navigation (tableau de bord, tableau multidimensionnels, graphes) complètent le panel des outils décisionnels.

Un système d'Information Géographique permet d'acquérir, de structurer, de mémoriser, d'analyser et de visualiser les données géographiques. Les utilisateurs potentiels d'un SIG sont tous les spécialistes qui ont besoin d'analyser d'importants volumes de données géographiques dans différents domaines.

Les Systèmes d'Aide à la Décision et les systèmes OLAP en particulier, ne présentent aucun instrument pour la gestion des données spatiales. Des solutions, connues sur le terme d'OLAP Spatial, qui visent à intégrer la donnée spatiale dans l'OLAP, ont donc été développées.

L'OLAP Spatial (SOLAP) a été défini par Yvan Bédard comme « *une plateforme visuelle conçue spécialement pour supporter une analyse spatio-temporelle rapide et efficace à travers une approche multidimensionnelle qui comprend des niveaux d'agrégation cartographiques, graphiques et tabulaires* » (Bédard, 1997). Le SOLAP enrichit les capacités d'analyse des systèmes OLAP classiques car la visualisation des mesures sur une carte permet de comprendre la distribution géographique d'un phénomène et de mettre en relation les différents phénomènes spatiaux par rapport aux axes d'analyse alphanumériques et de comparer ces phénomènes à diverses granularités géographiques. De plus, la composante cartographique dans l'OLAP représente une interface vers l'entrepôt de données spatiales. En d'autres termes, l'utilisateur peut accéder aux opérations de navigation multidimensionnelle à travers la simple interaction avec la composante cartographique.

L'intégration des données géographiques dans l'analyse en ligne est un enjeu majeur. La modélisation des entrepôts de données géographiques tout comme l'adaptation des fonctionnalités des systèmes d'entrepôt de données classiques pour les données géographiques, sont des problématiques ouvertes.

## Problématiques

L'information géographique est la représentation d'objets ou de phénomènes réels, localisés dans l'espace. Cette information est caractérisée par sa localisation dans l'espace, par sa forme et par ses aspects sémantiques, c'est à dire les attributs descriptifs alphanumériques et les relations avec d'autres objets.

Les modèles SOLAP existants se concentrent généralement sur la composante spatiale de l'information géographique. Ils définissent une mesure spatiale comme une collection d'objets spatiaux, et une dimension spatiale comme une dimension dont les membres contiennent un attribut spatial.

Ces modèles de données présentent une importante limite liée à la prise en compte de la composante sémantique de l'information géographique dans les mesures et dans les dimensions. D'un point de vue mesure, nous pensons que les attributs descriptifs des objets géographiques peuvent être utiles au processus décisionnel, pour expliquer un phénomène ou caractériser un ensemble de faits. Lorsque l'information spatiale est utilisée comme axe d'analyse, les modèles SOLAP existants utilisent les dimensions spatiales. Une dimension spatiale est décrite par des hiérarchies dont les membres sont des objets géographiques liés par relations topologiques d'inclusion ou d'intersection. Cette définition ne reflète pas la sémantique sous-jacente aux liens hiérarchiques. En effet, les objets géographiques peuvent être en relation avec d'autres objets à travers des relations spatiales, des relations de généralisation et des relations non spatiales. La prise en compte de ces types de relations est selon nous fondamentale dans l'analyse multidimensionnelle car à chaque type de hiérarchie correspond une analyse différente, qui peut se traduire en différentes politiques d'agrégation et de navigation.

Un deuxième point faible des solutions SOLAP existantes concerne les opérateurs d'analyse spatio-multidimensionnelle. Nous pensons que ces opérateurs devraient pouvoir modifier les dimensions géographiques selon l'exigence de l'utilisateur. En effet, au contraire de l'approche utilisée dans le processus décisionnel OLAP, l'analyse spatiale est flexible et itérative. Les données géographiques, grâce aux méthodes de transformations d'analyse spatiale, peuvent être modifiées ou remplacées tout au long du processus d'analyse. Or les opérateurs SIG fournis par les différents systèmes SOLAP sont des opérateurs orthogonaux aux opérateurs multidimensionnels, autrement dit, ils ne sont pas utilisés dans la navigation OLAP. Nous pensons donc que l'introduction et l'adaptation des opérateurs d'analyse spatiale dans un contexte OLAP afin d'intégrer la flexibilité de l'analyse spatiale au modèle multidimensionnel, est fondamentale pour une analyse spatio-multidimensionnelle satisfaisante. De plus, les modèles SOLAP existants utilisent des modèles de données différents pour la représentation des membres et des mesures spatiales. En termes d'analyse multidimensionnelle, cela se traduit par une asymétrie entre la mesure et la dimension spatiale, ce qui oblige l'utilisateur à concevoir des hypercubes différents pour changer le point de vue de l'analyse. Ce manque de symétrie et de flexibilité représente une limite importante des solutions SOLAP existantes, car l'information géographique doit pouvoir être aussi bien utilisée en dimension qu'en mesure. Enfin, une mesure en étant un objet géographique peut appartenir à des schémas hiérarchiques.

L'utilisation de cette information pour analyser les mesures à différentes granularités reste un défi important.

## **Contributions des nos travaux de recherche**

Nos travaux de recherche portent sur l'intégration de l'information géographique dans l'analyse multidimensionnelle.

Une analyse détaillée des concepts principaux du SOLAP, de l'information géographique et de l'analyse spatiale, nous a mené à la définition du concept d'OLAP Géographique (Tchounikine et al., 2005), (Bimonte et al., 2005), (Bimonte et al., 2006c), (Bimonte et al., 2007b). L'OLAP Géographique reformule les concepts du Spatial OLAP pour prendre en compte la composante spatiale et sémantique de l'information géographique et la flexibilité de l'analyse spatiale. L'OLAP Géographique définit les concepts de mesure, de dimension et de hiérarchie géographiques, ainsi que des nouveaux opérateurs multidimensionnels.

La mesure géographique étend le concept de mesure spatiale aux attributs alphanumériques d'un objet géographique, car ils sont nécessaires et complémentaires au processus d'analyse. Une dimension géographique est décrite par trois différents types de hiérarchies : descriptive, spatiale et de généralisation. Ces trois types de hiérarchies reflètent la sémantique des relations entre les membres de niveaux différentes. L'OLAP Géographique reformule également les opérateurs spatio-multidimensionnels, pour introduire dans un contexte multidimensionnel les opérateurs d'analyse spatiale, pour lever l'asymétrie entre mesure et dimension spatiale, et pour permettre d'exploiter les relations spatiales et aspatiales entre les mesures géographiques.

Nous proposons un modèle formel (GeoCube) (Bimonte et al., 2005), (Bimonte et al., 2006) et une algèbre associée qui prend en compte les concepts principaux de l'OLAP Géographique. L'originalité de GeoCube est la modélisation de toutes les données de l'univers d'analyse à travers les concepts d'objet complexe et/ou géographique. Les mesures, comme les membres de dimension, sont des objets géographiques décrits par un ensemble d'attributs descriptifs et un attribut spatial. La symétrie entre mesure et dimension se traduit notamment par l'appartenance des mesures à de schémas hiérarchiques. GeoCube, en s'appuyant sur cette modélisation des données particulière, propose une algèbre spatio-multidimensionnelle qui reformule et étend les opérateurs spatio-multidimensionnels classiques. L'algèbre fournit des opérateurs de forage et de coupe. L'opérateur de forage permet l'agrégation des objets géographiques, contrairement aux approches classiques où l'agrégation porte sur des mesures qui sont de simples valeurs quantitatives. L'opérateur de coupe permet de couper l'hypercube en utilisant des prédicats alphanumériques et aussi des prédicats spatiaux. Pour exploiter la symétrie entre mesures et dimensions, GeoCube introduit deux nouveaux opérateurs multidimensionnels qui permettent de naviguer dans la hiérarchie de la mesure et un opérateur qui permet d'invertir mesure et dimension Enfin, une autre innovation apportée par GeoCube est un opérateur qui permet de changer dynamiquement la structure de l'hypercube, en ajoutant dans les données décisionnelles les résultats obtenus par des opérateurs d'analyse spatiale.

Cette approche formelle a été implémentée dans le prototype GeWolap (Bimonte et al., 2006b), (Bimonte et al., 2007), (Bimonte et al., 2007b). GeWolap est une solution web OLAP-SIG intégrée qui gère les mesures géographiques et complexes, les dimensions géographiques, et implémente un ensemble d'opérateurs de l'algèbre de GeoCube. GeWolap permet une analyse spatio-multidimensionnelle satisfaisante grâce à une interface qui synchronise une composante tabulaire, une carte interactive et de affichages graphiques.

Enfin, nous nous sommes aussi intéressé à la visualisation des mesures géographiques dans un outil SOLAP. Nous avons proposé un nouveau paradigme de visualisation et d'interaction pour l'analyse des mesures géographiques (Bimonte et al., 2006c), (Bimonte et al., 2007a) qui combine les techniques OLAP classiques et une méthode de géovisualisation. En particulier, notre innovation consiste dans l'adaptation de la table de pivot et du Space-Time Cube pour l'analyse des données spatio-multidimensionnelles.

Notre travail de thèse s'inscrit dans le cadre d'une collaboration internationale avec l'organisation italienne CORILA (Consorzio per la Gestione del Centro di Coordinamento delle Attività di Ricerca inerenti il Sistema Lagunare di Venezia). Le but de cette organisation est la sauvegarde environnementale, architecturale et économique de la lagune de Venise. Notre équipe intervient dans l'axe de recherche « gestion des données ». L'objectif est de fournir aux spécialistes environnementaux un système pour l'analyse spatio-multidimensionnelle de données environnementales (Bimonte et al., 2006d). Ces données concernent la pollution des eaux de la lagune. Elles représentent les mesures de la pollution en 25 différentes zones de la lagune (appelées unités), à 2826 moments différents (3 années, 28 mois et 207 jours), et pour 100 polluants. Dans ce manuscrit, nous utilisons principalement les données environnementales de CORILA pour illustrer nos contributions.

## **Organisation du mémoire**

Pour présenter nos travaux, nous avons retenu pour ce mémoire de thèse une organisation en sept chapitres classés en trois parties.

La première partie est dédiée à l'état de l'art.

Le chapitre 1 introduit les principaux concepts des entrepôts de données et de l'analyse en ligne. Nous présentons les principes de la modélisation multidimensionnelle car nous nous appuyerons sur ces concepts pour la formalisation de notre modèle GeoCube. Le chapitre 1 présente également l'architecture classique des systèmes décisionnels.

Dans le chapitre 2, nous décrivons les caractéristiques fondamentales de l'information géographique en détaillant les aspects spatiaux et sémantiques qui la caractérisent. Ainsi, nous introduisons les concepts principaux des systèmes d'information géographiques : les bases de données spatiales, les méthodes d'analyse spatiales et les techniques de visualisation cartographique.

Le chapitre 3 introduit les concepts principaux de l'OLAP Spatial : mesures, faits et dimensions spatiales. Nous effectuons une classification de ces définitions et un état de l'art des différents modèles formels SOLAP. Nous présentons également un panorama des outils SOLAP existants.

La deuxième partie présente nos contributions.

Le chapitre 4 aborde une analyse critique des concepts du SOLAP par rapport à la prise en compte de la nature particulière de l'information géographique dans l'analyse multidimensionnelle. Cette analyse nous mène à la redéfinition des concepts du SOLAP. Nous définissons alors les concepts de mesure et de dimension géographique. Une dimension géographique est décrite par trois types possibles de hiérarchie : spatiale, descriptive et de généralisation.

Dans le chapitre 5, nous présentons notre modèle GeoCube et son algèbre. GeoCube formalise les concepts de l'OLAP Géographique présentés au chapitre 4. Le modèle est illustré en utilisant les données environnementales du projet CORILA.

Notre prototype GeWolap est présenté dans le chapitre 6. Une description de l'architecture du système précède la présentation des fonctionnalités de navigation et de visualisation de GeWolap. GeWolap gère les concepts de dimension et de mesure géographique, et implémente les opérateurs multidimensionnels définis par GeoCube. Grâce aux données concernant la pollution de la lagune de Venise, nous montrons comment ces nouveaux concepts sont pris en compte dans GeWolap, du point de vue de la modélisation logique et du point de vue de la visualisation. Un bilan critique de notre prototype qui met en évidence les points forts et faibles de notre solution conclut le chapitre.

Le chapitre 7 décrit le travail réalisé en collaboration avec Sergio Di Martino et Filomena Ferrucci de l'université de Salerno (Italie). Dans ce chapitre nous montrons que les techniques classiques de visualisation cartographiques ne sont pas adéquates pour l'analyse des mesures géographiques. En utilisant une maquette d'interface, nous proposons un nouveau paradigme d'interaction et de visualisation : la GeOlaPivot Table. A travers un cas d'étude concernant l'étude des maladies mortelles, nous décrivons notre technique de visualisation et d'interaction et ses caractéristiques principales.

La troisième partie présente un bilan de notre travail de thèse et les perspectives de recherche que nous envisageons pour compléter cette étude.



---

# Partie 1 : Etat de l'art

---





# *Chapitre 1 : Entrepôts de données et analyse en ligne*

Le but principal des Systèmes de Gestion de Bases de Données (SGBD) est de stocker, d'organiser et de permettre d'interroger au moyen de requêtes de grands volumes d'informations opérationnelles (Elmasri et Navathe, 2003). Ces données cachent souvent des connaissances et des informations stratégiques, toutes deux utiles au processus décisionnel.

Les Systèmes d'Aide à la Décision (SAD) sont des systèmes d'information flexibles et interactifs qui aident les décideurs dans l'extraction d'informations utiles pour identifier et résoudre des problèmes et pour prendre des décisions (Alter, 1980). Cette connaissance est obtenue à partir de données brutes, de connaissances personnelles et de modèles analytiques. Les SAD présentent ces informations, provenant de différentes sources, dans un environnement unique, uniforme et familier à l'utilisateur. Ils combinent, uniformisent et synchronisent les bases de données, les modèles d'analyse et les techniques de visualisation, en permettant de comparer différents résultats et de concevoir et valider des hypothèses. Parmi les systèmes d'aide à la décision, les systèmes d'entrepôts de données sont probablement les plus utilisés dans le monde académique et industriel.

Dans ce chapitre, nous décrivons les concepts principaux des entrepôts de données et de l'analyse multidimensionnelle et présentons les caractéristiques principales de leur modélisation formelle. Ensuite, nous décrivons les architectures des systèmes d'entrepôts de données.

## **1.1 Concepts principaux des entrepôts de données et de l'analyse en ligne**

Un entrepôt de données est « *une collection de données, intégrées, non volatiles et historisées pour la prise de décisions* » (Inmon, 1996). Dans un entrepôt de données, les données sont organisées en fonction des exigences analytiques des utilisateurs. Les entrepôts de données sont peuplés en utilisant différentes sources de données hétérogènes pour lesquelles ils fournissent une vision unifiée et homogène. Contrairement aux données opérationnelles qui ont une durée de vie limitée, les entrepôts de données doivent permettre une analyse historique. Les requêtes des bases de données opérationnelles s'effectuent sous forme de transactions qui lisent et écrivent un nombre réduit de lignes dans différentes tables liées par des liens référentiels. Ce type de requêtes est dit « On line Analytical Transactional Processing » (OLTP). Au contraire, le type de requêtes effectuées sur un entrepôt de données est appelé « On line Analytical Processing » (OLAP).

Les requêtes OLAP nécessitent la lecture d'une énorme quantité de données pour produire un ensemble de valeurs numériques. Les différences principales entre les systèmes OLTP et OLAP résumées par Kelly (Kelly, 1997) sont soulignées dans le Tableau 1.1.

	<b>OLTP</b>	<b>OLAP</b>
Utilisation	<i>Gestion des transactions</i>	<i>Aide à la Décision</i>
Conception	<i>Orientée applications</i>	<i>Orientée utilisateurs</i>
Fréquence	<i>Quotidienne</i>	<i>Sporadique</i>
Données	<i>Récents, détaillées</i>	<i>Historiques, multidimensionnelles, agrégées</i>
Source	<i>BD unique</i>	<i>Plusieurs BD</i>
Utilisation	<i>Répétitif</i>	<i>Ad hoc</i>
Accès	<i>Programmes précompilés</i>	<i>Outils d'Analyse</i>
Nb lignes accédées	<i>Dizaines</i>	<i>Milliers</i>
Type d'utilisateur	<i>Opérateurs</i>	<i>Décideurs</i>
Nb utilisateurs	<i>Milliers</i>	<i>Centaines</i>
Performance	<i>Elevée</i>	<i>Basse</i>
Dimension BD	<i>Giga-octets</i>	<i>Téraoctets</i>

Tableau 1.1. Différences principales entre les systèmes OLPT et OLAP (Kelly, 1997).

Les systèmes d'entrepôts de données reposent sur le paradigme multidimensionnel, qui grâce aux concepts de dimension, fait, mesure, hypercube et opérateurs OLAP, permet une analyse multidimensionnelle de grandes quantités de données.

### 1.1.1 Dimensions et hiérarchies

Les dimensions représentent les axes de l'analyse multidimensionnelle. Elles sont organisées en schémas hiérarchiques. Un schéma de hiérarchie, composé par plusieurs niveaux, représente différentes granularités ou degrés de précision de l'information. Un exemple de dimension peut être la dimension temporelle qui organise le temps grâce à une hiérarchie des niveaux jour, mois, années, ou encore une dimension représentant une classification de produits décrite par une hiérarchie avec les niveaux « Produit » et « Type » (Figure 1.1a). L'instance d'une dimension est un ensemble de membres. Ces membres sont connectés par des liens hiérarchiques en accord avec le schéma hiérarchique. Nous supposons que chaque hiérarchie contient un niveau (All), avec un seul membre. Un exemple d'une instance de la hiérarchie des produits est montré en Figure 1.1b, où par exemple les produits « Asp 321 » et « St 73 », appartenant au niveau « produit », sont liés au membre « Téléphone » du niveau « Type ».

Chaque niveau de la dimension peut présenter des attributs (Hüsemann et al., 2000) qui ne sont pas utilisées pour la définition du schéma hiérarchique, par exemple un produit peut présenter un attribut représentant, le prix, la couleur, etc. Ces attributs peuvent être utilisés dans l'analyse multidimensionnelle.

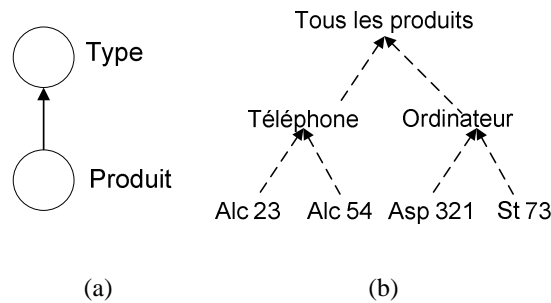


Figure 1.1. Hiérarchie des produits a) Schéma b) Instance.

Dans le monde réel, l'organisation hiérarchique des données est souvent très complexe. Différents auteurs ont proposé une classification de ces hiérarchies en utilisant la cardinalité des relations entre les membres de différents niveaux et les liens du schéma hiérarchique (Pedersen et al., 2001), (Malinowski et Zimányi, 2004a). Selon (Pedersen et al., 2001) une hiérarchie peut être *non-couvrante* si un membre « saute » un ou plusieurs niveaux de la hiérarchie, *non-onto* si un membre, non au niveau feuille, n'a pas de fils, et *non-stricte* si une relation plusieurs à plusieurs peut caractériser les membres de différents niveaux, par exemple un produit peut appartenir à plusieurs types. Enfin, l'expression *hiérarchies multiples* identifie le fait qu'une dimension peut être décrite par plusieurs hiérarchies, par exemple la dimension temporelle puisse être composée par une hiérarchie représentant le calendrier solaire et une autre associée au calendrier fiscal.

### 1.1.2 Faits et mesures

Un fait est un concept relevant du processus décisionnel et, typiquement, modélise un ensemble d'événements d'une organisation. Un fait est décrit par plusieurs mesures. Les mesures représentent usuellement des valeurs numériques qui fournissent une description quantitative du fait. Un fait est associé à une ou plusieurs combinaisons de membres des dimensions. Enfin, certaines mesures peuvent être calculées à partir d'autres mesures ou propriétés de membres. Elles sont appelées mesures dérivées (Blaschka et al., 1998).

Ainsi, une analyse multidimensionnelle portant sur un fait « ventes » d'un ensemble de magasins pourra être réalisée en définissant comme mesures « le volume des produits vendus », « le montant de la vente », et la mesure dérivée « profit », et comme dimensions « le temps », « les magasins », et « les produits » vendus. La Figure 1.2a représente le schéma de cette application grâce au modèle conceptuel multidimensionnel présenté dans (Malinowski et Zimányi, 2004a). Ce modèle permet d'examiner le volume et le montant totaux des produits vendus pour chaque mois et chaque année dans chaque magasin et dans chaque ville.

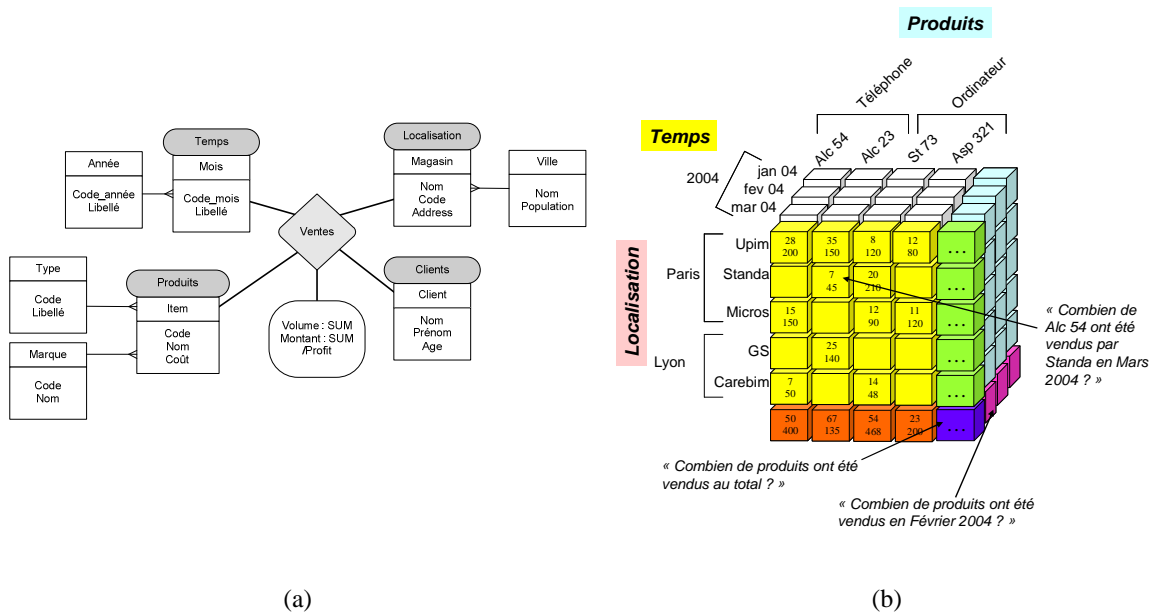


Figure 1.2. Application multidimensionnelle a) Schéma b) Hypercube.

A chaque combinaison des niveaux des dimensions correspond un niveau différent de détail des mesures. Dans les niveaux moins détaillés des dimensions les mesures sont agrégées en utilisant les fonctions d'agrégations. Les fonctions classiques pour agréger les mesures sont les mêmes que les opérations SQL "COUNT", "SUM", "MIN", "MAX" et "AVG".

Il est fondamental dans un modèle multidimensionnel de contrôler que le processus d'agrégation soit correct, afin de garantir une analyse multidimensionnelle exacte. Dans les systèmes OLAP, l'opérateur d'agrégation le plus utilisé est la somme. La précision de l'agrégation dépend de la sémantique de la mesure et de la structure de l'application multidimensionnelle. Par exemple, sommer des concentrations de pollution n'a pas de sens, en revanche le maximum est significatif. Il est possible de sommer la population des villes d'un pays, mais sommer la population d'une ville dans le temps est erroné, car les mêmes habitants seront comptés plusieurs fois. Cette problématique est connue dans l'OLAP sous le nom de problème d' « additivité » (Kimball, 1996). Une mesure est dite :

- *Additive* si l'on peut lui appliquer la somme sur toutes les dimensions.
- *Semi-additive* si la somme a du sens seulement sur certaines dimensions.
- *Non-additive* si elle n'est sommable sur aucune dimension.

Notons que la mesure peut être prise en compte plusieurs fois dans le processus d'agrégation, peut dépendre du type de hiérarchie, par exemple lorsque le modèle présente de hiérarchies non strictes ou multiples (Horner et al., 2004). Dans les bases de données statistiques, cette problématique est appelée « summarizability » (Lenz et Shoshani, 1997). La « summarizability » représente la situation dans laquelle le résultat d'une agrégation pouvait être calculé en utilisant les agrégations précédentes.

Une agrégation correcte implique :

- *Eviter de prendre en compte plusieurs fois la même mesure*
- *Respecter la sémantique de l'agrégation* : il s'agit de fournir un contrôle sur le type d'agrégation en prenant en compte la sémantique de la mesure.

### 1.1.3 Hypercube

L'instance d'un modèle conceptuel multidimensionnel est un hypercube. Un hypercube contient dans les cellules les valeurs des mesures détaillées et les axes sont faits par les membres des niveaux les plus détaillés de différentes dimensions. Ensuite, ce cube de base est décoré avec des cellules qui contiennent l'agrégation des valeurs de mesures pour chaque combinaison de membres des niveaux moins détaillés. Un exemple d'hypercube pour l'application de la Figure 1.2a est montré en Figure 1.2b. Sur les axes du cube de base on trouve les membres des niveaux des dimensions et dans les cellules les valeurs des deux mesures.

### 1.1.4 Requêtes multidimensionnelles

Le processus décisionnel multidimensionnel consiste en l'exploration de l'hypercube. L'utilisateur parcourt les données de l'hypercube selon les différents axes d'analyses à la recherche d'informations utiles, dans un processus fortement interactif, itératif et constructif, qui comprend des étapes de formulation des hypothèses, expérimentation et analyse (Tang et al., 2003). Les utilisateurs interagissent itérativement avec le modèle multidimensionnel pour formuler, modifier et valider leurs hypothèses. Les chemins d'analyse sont imprédictibles, contrairement aux données qui sont définies lors de la conception de l'application. Chaque résultat d'analyse est la conséquence des résultats précédents. Chaque étape du processus d'analyse est représentée par une navigation dans l'hypercube, ou par une requête multidimensionnelle. Ces requêtes utilisent les opérateurs OLAP. Les opérateurs OLAP permettent d'explorer les données multidimensionnelles en utilisant les concepts de dimensions et hiérarchies.

Un panorama des opérateurs OLAP proposés dans la littérature est présenté par Rafanelli en (Rafanelli, 2003). Les plus communs sont :

- *Les opérateurs de forage*
  - *Roll-up* permet de monter dans les hiérarchies des dimensions, et d'agréger les mesures.
  - *Drill-Down* est l'inverse du Roll-Up et permet de descendre dans une hiérarchie.
- *Les opérateurs de coupe*
  - *Slice* utilise un prédicat défini sur les membres des dimensions pour couper une partie de l'hypercube limitant le champ d'analyse et permettant à l'utilisateur de se concentrer sur des aspects particuliers du phénomène. En

utilisant la terminologie de l'algèbre relationnelle, l'opération de slice est l'équivalent de la sélection.

- *Dice* réduit la dimensionnalité de l'hypercube en éliminant une dimension. Cette opération est équivalente à la projection de l'algèbre relationnelle.
- *Drill-Accross* met en relation plusieurs hypercubes pour comparer leurs mesures. En effet dans un processus d'exploration et d'analyse, comparer plusieurs phénomènes est fondamental pour aboutir à une connaissance finale. Corréler plusieurs hypercubes pour avoir une vision unique des différentes mesures est donc nécessaire dans le processus d'analyse multidimensionnel. L'opération de drill-accross fusionne plusieurs hypercubes en utilisant les axes d'analyse en commun.

Un exemple de requête multidimensionnelle portant sur l'application de Figure 1.2a est : « *Quels sont le volume et le montant de chaque produit vendu par le magasin Carebim pour chaque année ?* ». Cette requête utilise à la fois l'opérateur de slice et celui de roll-up.

## **1.2 Les modèles conceptuels pour les bases de données multidimensionnelles**

La capacité de représenter l'information à deux niveaux différents et indépendants (abstrait et concret) est cruciale dans toutes les applications des systèmes d'information. Cette particularité est vraie aussi dans les entrepôts de données où le développement et la gestion des applications sont souvent source d'erreurs. Traditionnellement, les applications multidimensionnelles sont modélisées en utilisant des techniques très proches de l'implémentation. En général, l'utilisation d'un modèle logique est adaptée pour la description de l'implémentation dans un SGBD, mais il n'est pas assez expressif pour capturer efficacement les aspects essentiels de l'application multidimensionnelle. Il est reconnu que les modèles classiques, comme par exemple le modèle relationnel, ne sont pas adaptés à la modélisation des bases de données multidimensionnelles. Plusieurs efforts ont été réalisés ces dernières années en ce qui concerne la modélisation formelle de bases de données multidimensionnelles et des opérateurs OLAP. (Blaschka et al., 1998), (Pedersen et al., 2001) et (Torlone, 2003) décrivent les caractéristiques fondamentales d'un modèle conceptuel multidimensionnel.

Les propriétés fondamentales des dimensions comme les attributs de dimensions et la séparation explicite entre structure et instance sont prises en compte par quasiment tous les modèles, même si l'approche utilisée pour la modélisation des hiérarchies est très différente. Par exemple (Li et Wang, 1996) utilisent des relations de groupe, (Cabibbo et Torlone, 1998) des fonctions de roll-up, (Vassiadilis, 1998) et (Pedersen et al., 2001) des treillis de niveaux, (Trujillo, et al., 2000) des graphes acycliques, etc. La modélisation des hiérarchies complexes, aussi bien que les relations plusieurs à plusieurs (n-n) entre faits et dimensions a été abordée dans différents travaux, comme (Pedersen et al., 2001) et (Abello et al., 2006).

L'ensemble des modèles permet de représenter des ensembles de mesures. Par contre, comme pour le concept de hiérarchie, les approches utilisées sont différentes. Dans les modèles caractérisés par une symétrie entre dimensions et mesures, les mesures sont de simples valeurs alphanumériques indépendantes, de la même façon que les niveaux de dimensions (e.g. (Pedersen et al. 2001), (Agrawal, et al. 1997)). En revanche, dans les modèles où les mesures représentent explicitement et uniquement les valeurs descriptives d'un fait, celles-ci forment un ensemble dépendant de valeurs organisées dans une structure de données différente de celle utilisée pour représenter les membres de dimensions (e.g. (Trujillo et al., 2000), (Nguni et al., 2000)). Les mesures peuvent être des simples valeurs alphanumériques, d'ensembles de valeurs (e.g. (Vassiadilis, 1998), (Burdick et al., 2005), (Pedersen et al., 2001)) ou encore des mesures dérivées. Malgré l'importance de ce dernier concept dans les applications multidimensionnelles, peu de travaux se sont focalisés sur cette problématique car elle implique l'introduction d'un langage de calcul dans un contexte multidimensionnel (Blaschka et al., 1998).

Les modèles qui visent à garantir une agrégation correcte sont basés sur :

- La définition explicite du type de fonctions d'agrégations applicables aux mesures. Ces fonctions dépendent du type de mesure (e.g. (Pedersen et al. 2001), (Lehner, 1998)) et de la dimension sur laquelle on agrège (e.g. (Pedersen et al., 2002), (Abello, et al., 2006))
- Un contrôle qui évite de prendre en compte une mesure plusieurs fois, par exemple en prenant en considération les mesures détaillées (e.g. (Abello, et al., 2006), (Vassiadilis, 1998)), ou empêchant toutes formes d'agrégation si une agrégation n'est pas distributive (e.g. (Pedersen et al., 2001), (Pedersen et al., 2002))

Si différents modèles, que l'on peut qualifier de « généraux », ont été développés pour les bases de données multidimensionnelles classiques (Torlone, 2003), d'autres au contraire, s'attaquent à des problématiques très particulières. L'aspect incomplet de l'information est abordé par différents travaux, comme par exemple (Dyreson, 1996), (Pedersen et al., 2001) et (Ahmed et Miquel, 2005). (Pedersen et al., 2001), (Burdick et al., 2005) et (Delgado et al., 2004) se concentrent sur l'imprécision des informations en utilisant de fonctions particulières représentant le degré de précision ou la logique floue. L'aspect temporel est aussi à la base de nombreux travaux (e.g. (Body et al., 2003), (Eder et Koncilia, 2001), (Mendelzon et Vaisman, 2003), (Hurtado et al., 1999)). Les relations entre différentes structures de données multidimensionnelles et la navigation dans plusieurs hypercubes grâce à des opérateurs particuliers OLAP a été abordé dans différents travaux, comme (Schneider, 2003), (Abello et al., 2006) et (Teste, 2001).

Ces modèles révèlent le besoin et l'intérêt d'une solution formelle pour les systèmes d'entrepôts de données et OLAP. Si d'un côté ces solutions identifient très précisément les caractéristiques fondamentales d'un modèle formel, de l'autre, la prise en compte des aspects particuliers de l'information, comme l'imprécision, la composante spatiale, ou temporelle, ou encore les données multimédias, reste un défi majeur (Rizzi et al., 2006).

### 1.3 Architecture des systèmes d'entrepôts de données

Les architectures des systèmes d'entrepôts de données sont classiquement des architectures à trois niveaux, comme montré en Figure 1.3, constituées par un entrepôt de données, un serveur OLAP et un client OLAP.

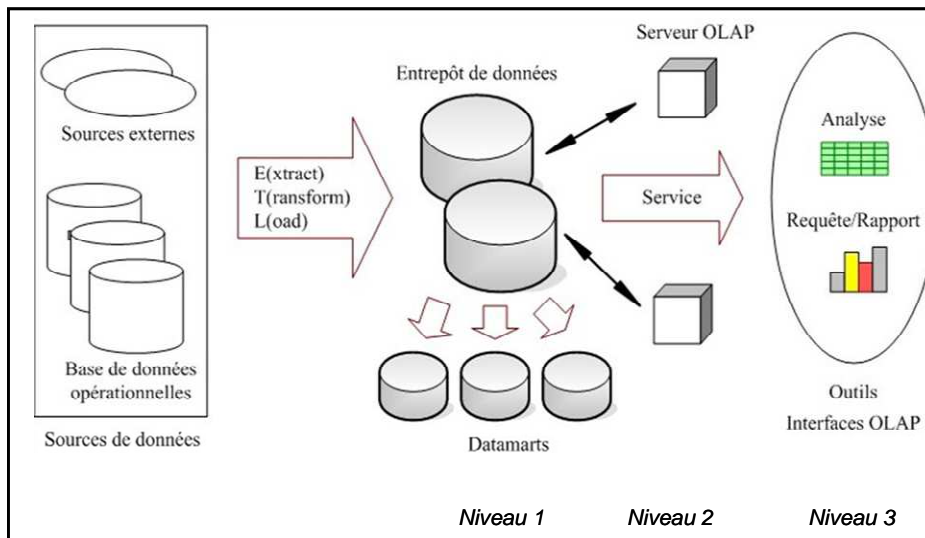


Figure 1.3. Architecture à trois niveaux d'un système d'entrepôt de données.

#### 1.3.1 Entrepôt de données

Le premier niveau est un SGBD. Les données sont extraites à partir des bases de données transactionnelles, nettoyées et transformées avec des outils ETL (Extract-Transform-Load ou en français extraction, transformation et alimentation), et intégrées dans l'entrepôt de données. Le SGBD contient aussi un ensemble de métadonnées concernant les sources de données, les mécanismes d'accès, les procédures de nettoyage et d'alimentation, les utilisateurs, etc.

#### 1.3.2 Serveur OLAP

Le deuxième niveau de l'architecture est un serveur OLAP, par exemple Mondrian, DB2 Olap Server, Oracle OLAP, Microsoft Analysis Services, etc. Le serveur OLAP permet d'effectuer une analyse de données conforme au paradigme multidimensionnel, avec des temps de réponse optimisés.

##### 1.3.2.1 Fonctionnalités d'un serveur OLAP

Un serveur OLAP fournit aux utilisateurs une vue multidimensionnelle des données qui peuvent être analysées grâce à un ensemble d'opérateurs OLAP (Roll-Up, Drill-Down, etc.). De plus, le serveur OLAP permet de gérer les données agrégées et les données détaillées de façon transparente. Les données d'un hypercube peuvent être représentées par un treillis hiérarchique de cuboïdes (Harinarayan et al., 1996). Le cuboïde le plus détaillé contient les valeurs des mesures détaillées. Les autres cuboïdes représentent l'agrégation de mesures détaillées. Si l'application présente  $n$  dimensions et chaque dimension  $L_i$  niveaux,



le nombre de cuboïdes est  $\prod_{i=1}^n L_i + 1$ . Le cuboïde avec les mesures aux niveaux d'agrégation les plus élevés est dit *cuboïde apex*. Le cuboïde qui présente les mesures détaillées est dit *cuboïde de base*.

Un exemple de treillis de cuboïdes utilisant les hiérarchies de Figure 1.4a et 1.4b est montré en Figure 1.4c.

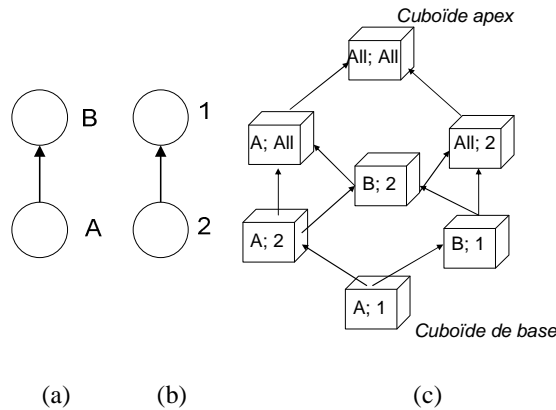


Figure 1.4. a) Hiérarchie avec les niveaux « A » et « B » b) Hiérarchie avec les niveaux « 1 » et « 2 » c) Treillis de cuboïdes.

Pour optimiser l'accès aux données agrégées, le serveur OLAP pré-calcule un sous-ensemble de cuboïdes. Lors d'une requête multidimensionnelle, le serveur OLAP cherche dans un premier temps si ces données ont déjà été calculées et si oui, il renvoie le résultat ; autrement il procède au calcul des cuboïdes nécessaires.

En plus des données agrégées et détaillées, un serveur OLAP gère aussi des métadonnées qui, de façon simple et transparente, masquent l'organisation multidimensionnelle complexe de l'entrepôt de données.

Les entrepôts de données contiennent souvent une grande quantité de données détaillées et agrégées, mais grâce à l'utilisation des techniques particulières d'indexation et de politiques de gestion des données pré-agrégées, le serveur permet le passage à l'échelle des données de l'application multidimensionnelle.

Un serveur OLAP ne se limite pas à fournir les opérateurs d'analyse multidimensionnelle, mais il met à disposition de l'utilisateur un puissant moteur de calcul qui présente des fonctions mathématiques et des procédures multidimensionnelles. Ces dernières permettent par exemple de naviguer dans les hiérarchies et de définir des mesures dérivées complexes.

Enfin, puisqu'un entrepôt de données est toujours analysé par plusieurs décideurs, un serveur doit gérer les accès concurrents, et garantir l'intégrité et la sécurité des informations.

### ***1.3.2.2 Implémentation physique d'un serveur OLAP***

Les approches principales pour l'implémentation de serveurs OLAP sont Multidimensional OLAP (MOLAP), Relational OLAP (ROLAP) et Hybrid OLAP (HOLAP).

Les serveurs MOLAP extraient les données de l'entrepôt de données et les mémorisent en utilisant des structures de données particulières et ils appliquent des techniques d'indexation et de hachage pour localiser les données lors de l'exécution des requêtes multidimensionnelles.

Les serveurs ROLAP (Kimball, 1996) utilisent la technologie des bases de données relationnelles pour mémoriser les données. Pour obtenir de performances acceptables, ces systèmes utilisent des structures d'indexation particulières comme l'index bitmap, et les vues matérialisées (Winter, 1998). Une vue matérialisée est la traduction relationnelle d'un cuboïde qui est pré-calculé et stocké dans l'entrepôt de données. La modélisation logique d'une base de données multidimensionnelles selon l'approche ROLAP ne systématise pas l'utilisation de la 3<sup>ème</sup> forme normale, contrairement aux systèmes OLTP. Le modèle logique le plus utilisé est le schéma en étoile (Kimball, 1996). Le schéma en étoile (Figure 1.5a) se constitue d'une table de faits et de tables de dimensions. Chaque entrée dans la table de faits représente un fait ou une cellule du cuboïde de base, et elle est liée, à travers des clés étrangères, à des dimensions. Les tables de dimensions sont des tables dénormalisées contiennent des attributs descriptifs et des attributs qui représentent les hiérarchies des dimensions. Par exemple, un produit est caractérisé par un code, par un nom, qui représente un attribut descriptif, et un type. Le type est utilisé pour regrouper les produits en catégories différentes. La dénormalisation améliore les performances des requêtes car elle réduit le nombre de jointures. Une variante du schéma en étoile est le schéma en flocon (Figure 1.5b). Il présente des dimensions partiellement ou totalement dénormalisées. La normalisation est utilisée pour réduire la redondance, surtout dans le cas où chaque niveau hiérarchique présente différents attributs et/ou quand la dimension est constituée de nombreux niveaux. Enfin, classiquement, un entrepôt de données est formé par différents hypercubes liés entre eux par des dimensions. La structure logique utilisée dans ce cas est le schéma en constellation. Il s'agit de plusieurs tables des faits qui partagent des tables des dimensions et qui peuvent être vu comme une collection d'étoiles (schéma en galaxie ou constellation de faits) (Figure 1.5c). Cette représentation logique permet d'utiliser l'opération de drill-across.

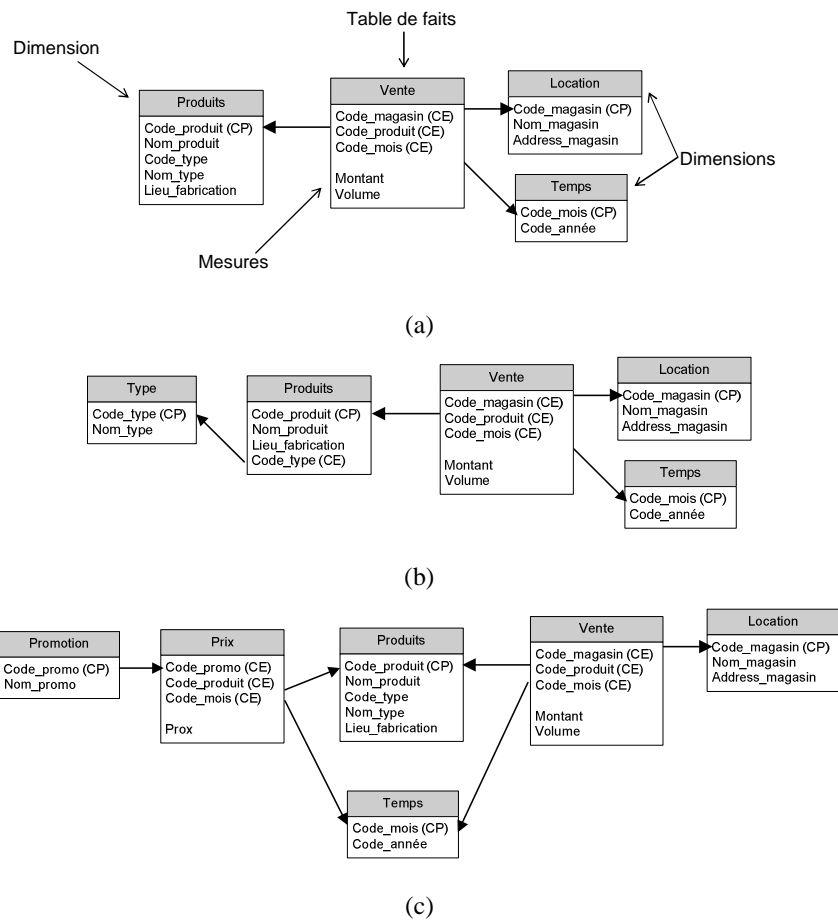


Figure 1.5. a) Schéma en étoile b) Schéma en flocon c) Schéma en constellation.

En général, les serveurs MOLAP (Thomsen, 1997) sont plus rapides et nécessitent moins d'espace de stockage, car les données sont stockées par le serveur OLAP de façon à garantir ces aspects. Les serveurs ROLAP permettent le passage à l'échelle et sont plus adaptés pour des mises à jour très fréquentes.

Enfin, il existe une autre typologie qui combine les deux technologies le Hybrid OLAP. Selon cette approche, une partie des données sont stockées en relationnel et l'autre en utilisant des techniques particulières.

### 1.3.3 Client OLAP

Le dernier niveau est un client OLAP, par exemple Microstrategy (Figure 1.6a), Cognos (Figure 1.6b), JPivot, Polaris (Stolte, et al., 2002), Advizor (Eick, 2000), etc., qui offrent une interface utilisateur avec des outils de reporting, d'analyse interactive, et parfois de fouille de données. Son rôle est de rendre l'information multidimensionnelle « visible », en d'autres termes, de permettre de découvrir des connaissances grâce à la seule visualisation et interaction avec les données. Parfois, le client et le serveur OLAP sont intégrés. Par exemple, Oracle (OracleBI, 2007) offre des outils d'analyse (Figure 1.6c) et des systèmes spécialisés pour une analyse facile des données issues des systèmes Oracle.

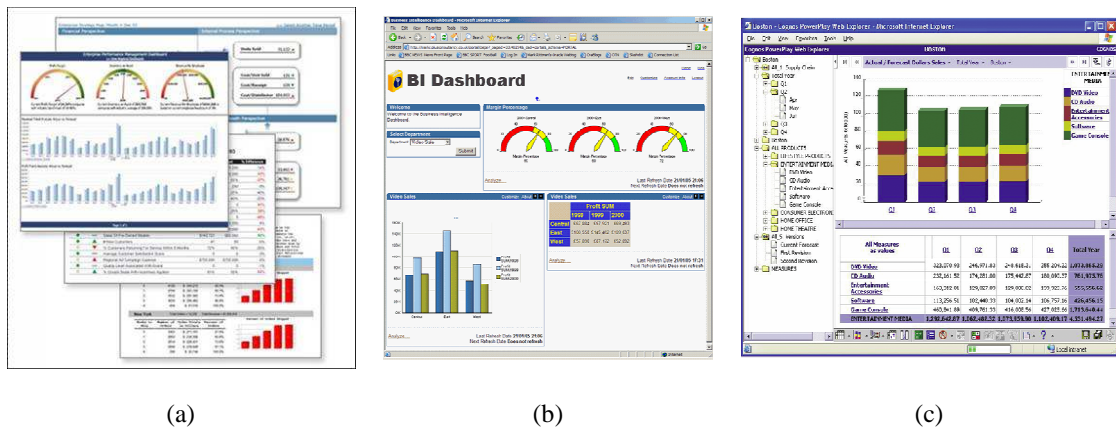


Figure 1.6. Clients OLAP a) Microstrategy b) Cognos c) Oracle.

Dans (Stolte et al., 2002), les auteurs définissent les caractéristiques nécessaires d'un outil d'analyse OLAP :

- *Affichage de grandes quantités de données* : Afficher de nombreuses mesures et dimensions.
- *Présence de multiples types d'affichage* : Fournir plusieurs types d'affichages graphiques (par exemple des histogrammes pour les ventes ou des couleurs pour les produits les plus vendus, etc.).
- *Emploi d'une interface pour l'exploration* : Comparer les données et changer les données et leurs modes de visualisation rapidement dans une interface familière et intuitive.

Le paradigme de visualisation le plus adopté par les clients OLAP est la table de pivot. Il s'agit d'un tableau multidimensionnel auquel sont associés les totaux et les sous-totaux, lequel tableau offre une vue de données imbriquées sur plusieurs niveaux. Le succès de la table de pivot est dû au fait qu'elle incorpore les concepts multidimensionnels (mesures, dimensions et hiérarchies) dans sa structure, générant un petit ensemble d'affichages, ce qui permet de comparer facilement et de façon interactive les données à différents niveaux de détail. Les interactions avec les différentes composantes correspondent aux opérateurs OLAP.

La Figure 1.7 montre un exemple d'une possible configuration de table de pivot pour l'application de Figure 1.2. Les trois mesures sont visualisées par magasin, ville et type de produits. Les cellules représentant les produits sont cliquables : un simple clic peut faire apparaître la liste de tous les produits pour chaque type. Il s'agit donc de la traduction visuelle de l'opérateur de Drill-Down.

Produits	Localisation	Mesures		
		Volume	Montant	Profit
All Produits	All Localisation	856,44	820,95	983,55
+Téléphone	-All Localisation	1 095,05	1 087,79	958,23
	+Paris	1 052,05	1 107,59	980,32
	-Lyon	1 088,31	884,90	1 085,73
	-Upim	1 336,43	943,27	1 117,62
	+Standa	969,34	1 086,70	1 094,71
	-Lille	964,30	900,58	953,65
	-GS	915,59	1 032,98	876,19
	+Micros	1 023,12	1 242,39	988,85
	-Nice	949,16	941,62	1 077,48
	-Carebim	1 055,92	1 095,56	1 078,26
	+Carefor	1 027,36	1 084,88	1 009,71
	+Grenoble	972,10	932,27	836,34
	+Nantes	919,44	926,47	873,20
	+Nice	968,02	1 174,67	830,33
+Ordinateur	-All Localisation	996,53	1 022,07	813,09
	+Paris	873,44	1 079,54	1 153,20
	-Lyon	1 116,23	1 010,00	759,70
	-Upim	921,89	938,78	888,19
	+Standa	965,60	1 065,26	810,71
	-Lille	907,71	973,44	1 079,99
	-GS	1 034,08	902,39	1 012,75
	+Micros	948,53	985,58	1 144,74
	-Nice	987,59	1 067,84	1 070,92
	-Carebim	876,23	1 049,53	890,92
	+Carefor	1 106,05	855,01	1 122,72
	+Grenoble	1 037,09	1 073,14	1 044,79
	+Nantes	958,40	997,69	823,23
	+Nice	1 047,98	1 073,05	1 028,30

Figure 1.7. Visualisation des données de l'application multidimensionnelle des ventes de Figure 1.2 avec la table de pivot de JPivot.

Les tables de pivot sont souvent couplées avec des affichages graphiques, par exemple de barres 3D, de plots à bulles, etc., qui permettent de « comprendre » visuellement les données. Dans un client OLAP, toutes ces composantes sont synchronisées au sens qu'une action (par exemple un drill-down) sur une composante (par exemple la table de pivot) est propagée aux autres composantes (les affichages graphiques).

En conclusion, le client OLAP permet aux utilisateurs d'explorer itérativement les données en accord avec le paradigme multidimensionnel. Le caractère extemporané des sessions de travail OLAP, la forte dépendance des données, la complexité des requêtes et l'orientation aux utilisateurs non informaticiens, rendent cet instrument nécessaire et fondamental dans le processus décisionnel.

## 1.4 Conclusions

Dans ce chapitre, nous avons présenté un panorama des concepts principaux des systèmes d'entrepôts de données et OLAP. Nous avons décrit les concepts à la base de l'analyse multidimensionnelle, structures de données et opérateurs, et nous avons détaillé les trois niveaux de l'architecture classique des systèmes d'entrepôts de données. Nous avons également présenté les caractéristiques fondamentales des modèles formels pour les bases de données multidimensionnelles, qui sont centrées autour d'un ensemble bien défini des caractéristiques fondamentales. En revanche, lorsque des aspects particuliers, comme l'imprécision ou l'incomplétude, sont introduits dans l'information multidimensionnelle, une reformulation des approches formelles OLAP devient nécessaire tout comme une révision des fonctionnalités des différents niveaux de l'architecture classique. C'est dans cette direction que se positionnent les travaux concernant l'introduction de l'information spatiale dans l'OLAP, comme décrit dans les chapitres suivants.



# *Chapitre 2 : L'information géographique et les Systèmes d'Information Géographique*

Un système d'Information Géographique (SIG) est un ensemble d'outils qui permet d'intégrer des informations cartographiques et alphanumériques dans un environnement unique et cohérent. Les SIG permettent d'accomplir un ensemble d'opérations liées aux données géographiques comme l'acquisition, la structuration, la mémorisation, l'analyse et la visualisation. Les utilisateurs potentiels d'un SIG sont tous les spécialistes qui ont besoin d'analyser d'importants volumes de données géographiques dans différents domaines (urbanisme, environnement, climatologie, marketing, agriculture, santé, transport, etc.).

Dans ce chapitre, nous décrivons tout d'abord les caractéristiques principales de l'information géographique. Ensuite, nous nous focalisons sur les fonctionnalités des systèmes d'information géographique, par rapport à la mémorisation, à l'analyse et à la visualisation de données géographiques.

## **2.1 Les données géographiques**

L'information géographique est la représentation d'un objet ou d'un phénomène réel localisé dans l'espace à un moment donné. L'information géographique est caractérisée par une composante purement spatiale et une composante sémantique (Degréne et Salgé, 1997).

### **2.1.1 La composante spatiale**

La composante spatiale représente la position sur la surface terrestre et la forme d'un objet du monde réel. Une position est décrite dans un système de référence explicite comme par exemple un système de coordonnées. Cette composante permet de représenter la forme de l'objet lui-même et de positionner celui-ci par rapport aux autres phénomènes ou objets du monde réel. En Figure 2.1 est montrée la composante spatiale de l'information géographique représentant les départements de France. Chaque département possède une forme et occupe une position précise par rapport aux autres départements.

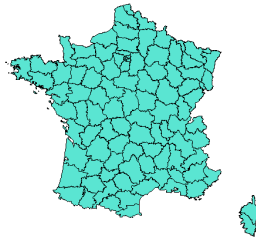


Figure 2.1. La composante spatiale des départements de France.

### 2.1.2 La composante sémantique

La composante sémantique représente l'information relative à la nature, l'aspect et les propriétés descriptives d'un objet ou à un phénomène du monde terrestre, par exemple un département est décrit par son nom (Rhône), sa population (1508966), etc.

Cette information peut aussi inclure des relations avec d'autres objets ou phénomènes, par exemple le département Rhône appartient à la région Rhône-Alpes. Un des aspects sémantiques qui distingue l'information géographique des données classiques est sa représentation multiple à différentes échelles (Laurini et Milleret-Raffort, 1993), (Weibel et Dutton, 1999), que nous détaillons ci-dessous.

### 2.1.3 La généralisation

Dans le domaine de la cartographie traditionnelle, la généralisation est le processus de réduction de complexité d'une carte dans le processus de réduction d'échelle, en gardant l'essentiel et en supprimant ce qui n'est pas important, en maintenant les relations locales et non ambiguës entre les objets géographiques et une représentation cartographique claire. La généralisation des cartes est « *la sélection et la représentation simplifiée à un niveau de détail approprié à l'échelle et/ou à l'objectif de la carte* » (ICA, 1973).

La réduction de l'échelle implique une compétition pour l'espace parmi les objets géographiques d'une carte. Des ensembles d'objets géographiques sont éliminés ou remplacés par des objets de taille différentes, ou par des symboles, d'autres encore sont déplacés, leurs formes sont simplifiées et les relations topologiques entre eux peuvent être changées. Cependant, l'échelle n'est pas le seul facteur qui influence la généralisation. L'objectif de la carte joue aussi un rôle très important. Par exemple une carte pour les cyclistes mettra en évidence des sections de routes différentes de celles qui peuvent être utilisées par les voitures. Un exemple classique de généralisation est la dérivation d'une carte topographique d'une échelle 1:100000 à une échelle 1:50000.

Dans les systèmes de cartographie numérique et dans les SIG, la généralisation est « *le processus qui réalise une transition entre différents modèles représentant une partie du monde réel à un niveau diminuant de détail, tout en maximisant le contenu informatif par rapport à une application donnée* » (Weibel et Dutton, 1999). La généralisation permet de dériver des données et des cartes pour des échelles secondaires et/ou thèmes particuliers



tout en gardant une bonne lisibilité cartographique et une représentation focalisée autour du thème principale de la carte.

(McMaster et Shea, 1992) proposent un cadre conceptuel pour la généralisation des données géographiques, qui comprend trois phases : la première prend en compte l'objectif de la généralisation (i.e. l'objectif de la carte, etc.) (pourquoi), la deuxième concerne une évaluation cartométrique des conditions qui indiquent quand généraliser (quand) et la troisième est la sélection appropriée des attributs spatiaux et alphanumériques et leur transformation à travers un ensemble d'opérateurs (Regnauld et McMaster, 2007) (comment). Ces opérateurs établissent des relations complexes entre les objets géographiques de la carte originale et de la carte généralisée (Neun et Steiniger, 2005), (Bobzien et al., 2006).

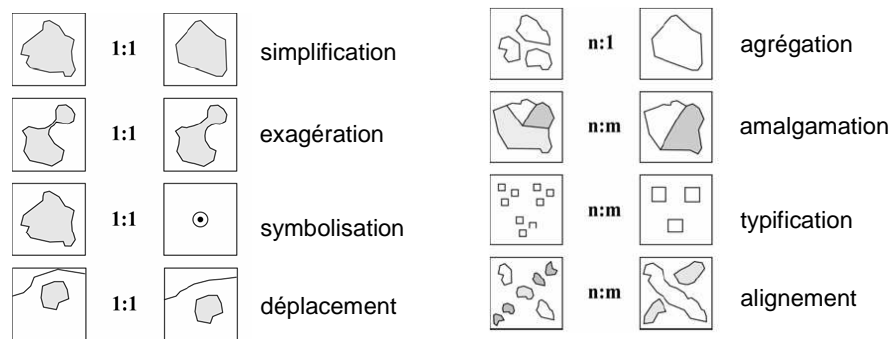


Figure 2.2. Exemples de relations d'identité et de groupe (Neun et Steiniger, 2005).

Les relations principales entre ces objets sont les relations d'identité (1 : 1) qui peuvent exister entre deux objets de deux différentes cartes et les relations de groupe (n : m) (Figure 2.2). De plus, le processus de généralisation peut créer des relations 1 : 0 comme conséquence des opérateurs de sélection ou abstraction qui éliminent un objet de la carte originale. Une organisation hiérarchique de l'information géographique a été utilisée dans différents travaux. (Timpf et Frank, 1995) proposent l'utilisation de graphes acycliques directs pour représenter les relations entre les différentes cartes. (Stell et Worboys, 1998) organisent la base de données spatiale comme une carte stratifiée de l'espace. Une carte stratifiée est un ensemble de cartes et grâce à l'utilisation d'opérations particulières de transformation l'utilisateur peut naviguer dans ces cartes. (Timpf, 1998) représente cette hiérarchie grâce à la combinaison de trois types différents de hiérarchie : généralisation, agrégation et filtrage.

Ces relations, qui mettent en correspondance les objets d'une carte avec la même information du monde réel mais avec un autre niveau de détail, sont utiles pour différentes tâches, par exemple pour la mise à jour d'une base de données géographiques, ou l'affichage des cartes sur des dispositifs particuliers, et aussi dans le processus décisionnel spatial (Vangenot, 2001). La généralisation fournit aux décisionnaires une vision à différents niveaux de détail du contexte ou du phénomène géographique, en permettant de se focaliser seulement sur les aspects décisionnels les plus importants et en les stimulant dans la formulation et la vérification d'hypothèses.

## 2.2 Fonctionnalités des Systèmes d'Information Géographique

Différentes solutions SIG commerciales et non commerciales ont été développées (ArcView, MapInfo, Quantum GIS, Grass, etc.). Dans les SIG les données géographiques, organisées sous forme de couches, sont mémorisées selon deux différents modèles.

### 2.2.1 Les modèles matriciel et vectoriel, et les couches

L'information géographique est représentée à travers deux types de modèles ou structures de données : les modèles vectoriel et matriciel (Rigaux, et al., 2001). Le choix dépend de la nature des données géographiques et de leur utilisation (Couclelis, 1992). Le modèle matriciel (dit aussi Raster) représente l'espace comme un seul champ continu. Au contraire, le modèle vectoriel modélise l'espace à travers ensemble d'objets identifiés grâce à un ensemble de points.

Le modèle matriciel permet de manipuler et de représenter l'information cartographique à partir d'une matrice de cellules (pixels) de même forme et de même dimension qui possède certains attributs. Cette représentation est la plus adaptée pour la modélisation des phénomènes purement spatiaux (e.g. la température, la pollution, etc.), mais étant donné que les pixels de la matrice n'ont pas de signification d'ensemble, seule une analyse thématique effectuée grâce à la « Map Algebra » (Tomlin, 1990) est possible. La « Map Algebra » est l'algèbre des cartes. Elle permet de créer des informations spatiales et thématiques nouvelles sous forme d'une matrice. Ces informations sont le(s) résultat(s) de calculs réalisés à partir d'une ou plusieurs matrices et ceci maille par maille, ou zone par zone.

La structure vectorielle permet de manipuler et de représenter les données géographiques d'après les coordonnées de points individuels auxquels on peut ajouter des attributs. Dans le modèle vectoriel, à chaque forme géométrique (point, ligne, polygone, etc.) sont associées ses informations descriptives en utilisant une autre structure de données. Ce modèle est adéquat pour la représentation de l'information géographique défini par les hommes, par exemple des bâtiments ou les limites administratives d'un pays. La caractéristique principale de cette approche est la représentation géométrique des entités géographiques selon l'approche de la cartographie traditionnelle. Le modèle vectoriel permet de représenter explicitement les frontières d'une géométrie autorisant l'analyse des objets géographiques grâce à leurs relations spatiales, soit topologiques (à l'intérieur, à l'extérieur, etc.), de distance et directionnelles (nord, sud, etc.) (Longley, et al., 2001), et grâce à leurs propriétés métriques (la surface, etc.). Plusieurs auteurs adoptent le modèle vectoriel pour définir l'information géographique. (Rigaux et al., 2001) définit un objet géographique comme une entité du monde réel décrite par deux composantes : la composante descriptive qui est un ensemble d'attributs alphanumériques (par exemple le nom, la population etc.) et la composante spatiale qui correspond à sa géométrie et ses relations topologiques. (Worboys, 2004) appelle un objet avec référence spatiale ou géo-objet, une entité avec des attributs spatiaux, temporels, graphiques et alphanumériques.

La métaphore utilisée dans le SIG pour organiser les données géographiques est la couche. Une couche est une carte qui représente un ensemble d'informations géographiques du même type (routes, bâtiments, cours d'eau, limites de communes, la population, le type

de sol, etc.) (Figure 2.3). Cette structuration rend ces informations superposables, compatibles, combinables et par conséquent analysables.

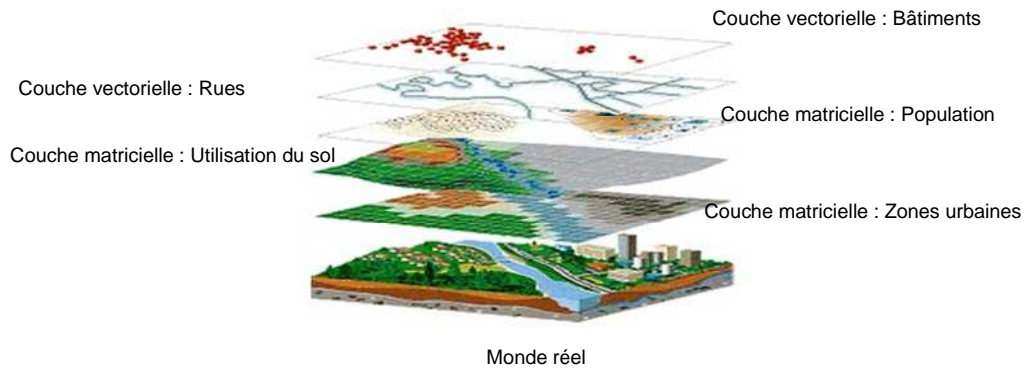


Figure 2.3. Organisation des informations géographiques en couches.

### 2.2.2 Les bases de données spatiales

Une fonctionnalité des SIG est d'intégrer des données de sources différentes dans un cadre conceptuel et logiciel commun. Pour la mémorisation de l'information géographique associée aux différentes couches deux approches différentes sont utilisées : faiblement ou fortement intégrées (Rigaux et al., 2001). Dans un SIG fortement intégré les informations spatiales et alphanumériques sont stockées dans un seul système. La donnée spatiale est mémorisée dans le même SGBD que celui utilisé pour les données alphanumériques (SGBD spatial). Les SGBD spatiaux (Oracle Spatial, PostGIS, SpatialWare for SQL Server, DB2 spatial extender, etc.) étendent les SGBD classiques avec des types de données spatiales (e.g. point, ligne et polygone) et avec des fonctions spatiales (e.g. overlap, distance, et surface). Ils permettent de définir un attribut d'une table de type spatiale, formuler des requêtes qui utilisent de prédicats spatiaux, agréger les données spatiales, améliorer les temps de réponse grâce à des index spatiaux, etc. Toutes ces fonctionnalités sont implémentées d'après des standards existants, comme SFS pour SQL (OpenGIS, 1999) ou SQL/MM (ISO, 2002) qui définissent une extension du SQL classique pour les données spatiales. Si d'un côté le modèle vectoriel est bien implémenté dans les SGBD spatiaux existants, de l'autre, un petit nombre de systèmes (e.g. (Oracle10g, 2007)) gère les données géographiques modélisées en format matriciel. L'approche fortement intégrée permet de mémoriser les données spatiales avec les données sensibles de l'application en diminuant la complexité du système, et autorise un accès standard aux informations spatiales. De plus, les SGBD spatiaux permettent aux applications SIG de s'enrichir des caractéristiques principales des bases de données : sécurité, concurrence d'accès, flexibilité et passage à l'échelle.

Un exemple de création d'une table spatiale représentant les départements de France à travers le modèle vectoriel et d'une requête spatiale qui vise à sélectionner les départements avec une surface supérieure à 5000 Km<sup>2</sup> en utilisant l'extension d'Oracle pour les données spatiales (OracleSpatial, 2007) sont montrés respectivement en Figure 2.4a et 2.4b. L'insertion d'un tuple dans la table et sa représentation cartographique sont montrées en figures 2.4c et 2.4d.

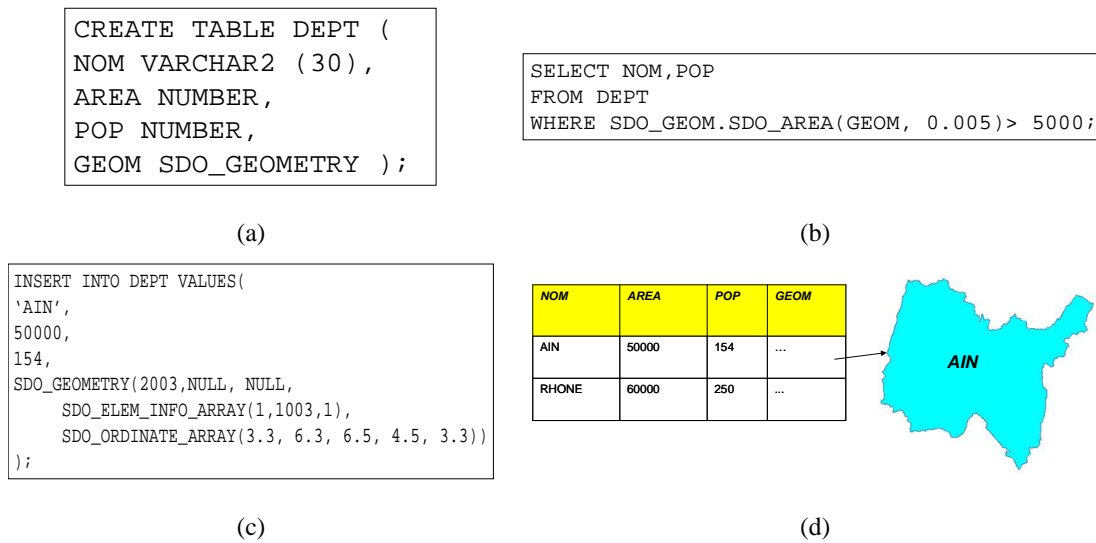


Figure 2.4. Oracle Spatial a) Création d'une table spatiale b) Requête SQL qui utilise en prédicat spatial c) Insertion d'un objet géographique dans la table d) Représentation tabulaire et cartographique.

Dans l'approche faiblement intégrée, les applications SIG stockent les attributs spatiaux et alphanumériques séparément. Les représentations des objets spatiaux sont gérées par des systèmes de gestion de fichiers, alors que la composante descriptive est mémorisée dans des SGBD classiques.

### 2.2.3 L'analyse spatiale

Parmi les fonctionnalités des SIG, les opérateurs d'analyse spatiale jouent un rôle très important. Le terme d'analyse spatiale fait référence à l'analyse d'un phénomène distribué dans l'espace qui possède également des dimensions physiques (la localisation, la proximité, l'orientation, etc.). Elle vise à l'estimation, la prédiction, l'interprétation et la compréhension des phénomènes du monde réel, en mettant en évidence des structures et des formes d'organisation spatiale récurrentes. Elle s'applique à plusieurs domaines comme les problèmes environnementaux, le marketing, les sciences humaines, etc.

Comme toute discipline scientifique, l'analyse spatiale rassemble différentes méthodes, outils et techniques. Pour l'analyse spatiale, les utilisateurs des SIG ont à disposition un ensemble d'outils statistiques (par exemple des méthodes de Point Pattern Analysis) et non statistiques (par exemple la détermination des zones tampon).

Les méthodes d'analyse spatiale ont été classifiées par (Longley et al., 2001) de la façon suivante :

- *Les méthodes d'interrogation et raisonnement* représentent les formes les plus simples d'analyse spatiale. Le SIG est utilisé pour répondre aux questions sous la forme de requêtes qui utilisent des prédicats spatiaux ou alphanumériques pour sélectionner un ensemble d'objets géographiques. Aucun changement

n'intervient dans les bases de données spatiales. La requête : « *Quelles sont les régions de France traversées par le Rhône ?* » peut être un exemple.

- *Les méthodes de mesure* permettent d'obtenir des valeurs numériques qui décrivent des objets géographiques. Elles incluent des mesures comme le périmètre, la surface pour un seul objet spatial et la distance entre deux objets spatiaux.
- *Les méthodes de transformation* sont des opérations spatiales qui changent les données en les combinant entre elles pour obtenir de nouvelles données géographiques. Ces transformations utilisent de simples règles géométriques, logiques et/ou arithmétiques, par exemple l'overlay, la fusion, etc.
- *Les méthodes de synthèse* synthétisent un ensemble de données en une ou deux valeurs numériques. Elles sont les équivalents spatiaux des descripteurs statistiques classiques et elles incluent des méthodes d'analyse statistiques spatiale, comme la « standard distance » ou le barycentre.
- *Les techniques d'optimisation* sont utilisées pour sélectionner la position idéale pour des objets géographiques selon des critères bien définis.

Le processus d'analyse spatiale est souvent associé à une simple séquence d'opérations de la spécification du problème au résultat. Dans la réalité, cette définition simplifie énormément le processus d'analyse, car comme il a été montré par différents travaux, il est beaucoup plus complexe et itératif. (Mitchell, 2005) formule le processus d'analyse spatiale comme un processus itératif (Figure 2.5) qui consiste en :

- Identifier le problème et les buts de l'analyse,
- identifier les problématiques spatiales et les outils pour les résoudre,
- identifier les données et leur préparation pour les opérations spatiales,
- créer un plan d'analyse (les fonctions et les opérations spatiales peuvent être appliquées séquentiellement, en d'autres termes leur résultat peut être l'entrée d'une ou plusieurs fonctions et/ou opérations spatiales),
- exécuter le plan, et
- visualiser les résultats.

(Haining, 2003) étend cette définition, permettant à chaque étape du processus d'analyse de pouvoir retourner aux étapes précédentes et réévaluer les décisions prises à ces points.

En d'autres termes, le processus d'analyse spatiale est flexible et itératif et les connaissances de l'expert du domaine sont fondamentales dans la reformulation de différentes étapes pour aboutir à une connaissance satisfaisante.

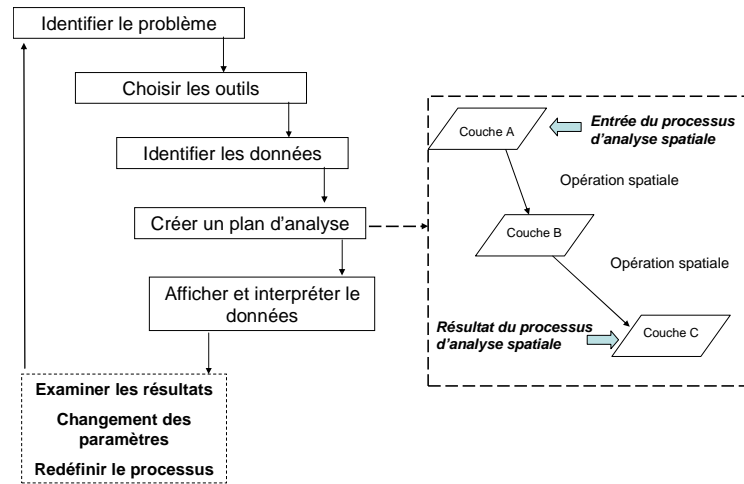


Figure 2.5. Processus d'analyse spatiale (Mitchell, 2005).

## 2.2.4 De la visualisation à la géovisualisation

Dans le processus d'analyse spatiale, la visualisation joue un rôle très important. Les cartes ne visualisent pas seulement les données géographiques, mais elles stimulent également l'utilisateur dans son processus mental pour la découverte de patterns, relations et tendances (MacEachren, 2001).

### 2.2.4.1 La visualisation cartographique

L'information géographique peut être transmise de manière verbale ou chiffrée, graphique ou cartographique. Les représentations visuelles sont particulièrement efficaces, parce qu'elles impliquent au préalable un traitement des données et une sélection de l'information pertinente et parce qu'elles autorisent une lecture instantanée d'une image globale. La transmission visuelle n'est réalisée correctement que si le codage par les signes graphiques a été opéré selon les règles assez strictes. Ces règles ont été dictées à partir de recherches sur la cognition : comment l'œil humain perçoit-il les signes graphiques et comment le cerveau transcrit-il cette perception en information ? Pour réaliser une représentation cartographique, le cartographe dispose de signes élémentaires, qu'il distribue selon une certaine implantation graphique. Il combine ces signes, pour former des figurés, en fonction de six variables visuelles (Bertin et Bonin, 1992). La graphique utilise les propriétés du plan pour faire apparaître les relations de ressemblance, d'ordre ou de proportionnalité entre les ensembles de données. La sémiologie graphique est l'ensemble des règles qui permettent l'usage d'un système de signes graphiques pour transmettre l'information. Le langage cartographique en est une partie. Les signes graphiques élémentaires (points, lignes, surfaces) sont l'alphabet, les variables visuelles sont le vocabulaire, et les règles de la perception visuelle sont la syntaxe. On appelle variable visuelle une façon de faire varier les signes graphiques. On distingue six variables visuelles différentes : la taille, la valeur, la couleur, la forme, l'orientation et le grain (Tableau 2.1).

Le figuré cartographique est construit par le cartographe à partir des signes graphiques élémentaires. C'est une construction qui peut recevoir des implantations graphiques différentes et qu'on peut faire varier en utilisant les variables visuelles. L'implantation

graphique est la manière d'appliquer le figuré sur la carte : elle peut être ponctuelle lorsque le figuré est attribué à un point ou à un symbole linéaire lorsque le figuré est affecté à une ligne, et zonale lorsque le figuré est étendu sur une surface.

La visualisation de données géographiques est un problème complexe. En effet, pour permettre une analyse judicieuse des cartes, il faut savoir trouver un juste équilibre entre la précision (et donc le nombre de données représentées sur la carte) et la lisibilité. En général, les SIG mettent à disposition de l'utilisateur, deux types différents de cartes : les cartes de références générales et les cartes thématiques. Les cartes de références générales représentent la localisation d'objets concrets (cartes routières, etc.). Les cartes thématiques représentent la distribution spatiale d'attributs (par exemple des cartes représentant la densité de population, etc.), grâce aux variables visuelles et aux affichages graphiques (bars, camemberts, etc.)

Dans les SIG les informations géographiques affichées sur les cartes peuvent être triées, sélectionnées et traitées en fonction des demandes. Il est ainsi possible de présenter les données selon un ou plusieurs thèmes, de les faire apparaître selon l'échelle de visualisation et de limiter leur usage en fonction de leur précision et de leur accessibilité. De plus, l'utilisateur peut à tout moment avoir accès aux données alphanumériques associées à la composante spatiale à travers des requêtes à la base de données, des outils graphiques et surtout à travers l'interaction avec la carte. Cette capacité confère une nouvelle perspective aux cartes en les transformant en de véritables outils d'analyse et d'exploration. La carte peut être considérée comme une interface vers l'information géographique. Elle permet d'explorer les données géographiques en gardant en même temps son rôle premier : la présentation de données dans l'espace.


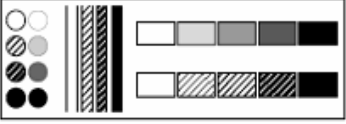
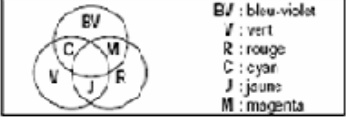
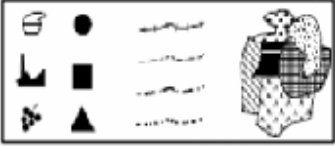

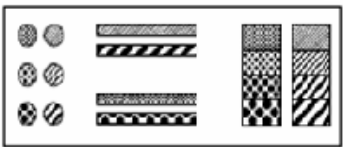
	<p><b>La taille</b></p> <p>La variation de taille permet de traduire parfaitement la variation quantitative.</p>
	<p><b>La valeur</b></p> <p>La variation de valeur d'une couleur est une variation d'intensité lumineuse du plus sombre au plus clair, ou inversement; elle traduit une relation d'ordre et des différences relatives (relation quantitative)</p>
	<p><b>La couleur</b></p> <p>Comme la taille, les couleurs traduisent des différences mais qui ne peuvent cependant être ordonnées entre elles. Elles sont le plus souvent chargées de significations culturelles et psychologiques (bleu pour l'eau).</p>
	<p><b>La forme</b></p> <p>La forme se définit comme la variation du contour d'une tache. Elle cherche à exprimer l'identité de l'objet à représenter. Il peut s'agir de formes géométriques fondamentales (carré, cercle, ...) ou de pictogrammes.</p>
	<p><b>L'orientation</b></p> <p>L'orientation permet de positionner un signe par rapport aux deux axes du graphique; elle exprime les différences et gagne en efficacité en combinant les variables de grains et de valeur</p>
	<p><b>Le grain</b></p> <p>Les grains constitutifs des trames combinent déjà plusieurs variables (forme, taille) et traduisent une relation d'ordre et des différences relatives (relation quantitative)</p>

Tableau 2.1. Variables visuelles (Bertin et Bonin, 1992).

#### 2.2.4.2 La géovisualisation

Dans les années 90 la visualisation scientifique donnait au mot « visualisation » une nouvelle signification en liant la visualisation à l'informatique décisionnelle. Autrement dit, la visualisation scientifique devenait de plus en plus un outil d'aide à la décision permettant de rendre visibles les connaissances qui se cachent derrière les données.

Les relations entre les domaines de la cartographie et du SIG d'un côté, et la visualisation scientifique de l'autre, sont à la base d'une nouvelle discipline scientifique : la Géovisualisation. Celle-ci intègre les techniques de visualisation scientifique, de cartographie, d'analyse des images, d'exploration de données, pour fournir une théorie, des méthodes et des outils pour la représentation et la découverte de la connaissance spatiale



(MacEachren et Kraak, 2001). Dans ce contexte, il est fondamental que les procédures de cartographie fassent plus attention aux interfaces homme-machine et donc aux problématiques de représentation et de calcul dans le processus de visualisation et aux capacités de facilité d'utilisation et surtout d'interaction.

Ce lien indissoluble entre l'interactivité, et l'exploration et l'analyse de données a été bien mis en évidence par (MacEachren et Kraak, 1997) à travers le « map-use cube » (Figure 2.6). Le « map-use cube » représente visuellement le degré d'interaction (« Interaction »), le type de cible (« Audience ») et le degré de connaissance des données (« Data Relations ») dont l'utilisateur a besoin dans les étapes d'exploration (« Explore »), d'analyse de données (« Analyse »), de synthèse (« Synthesize ») et de présentation des résultats (« Present »).

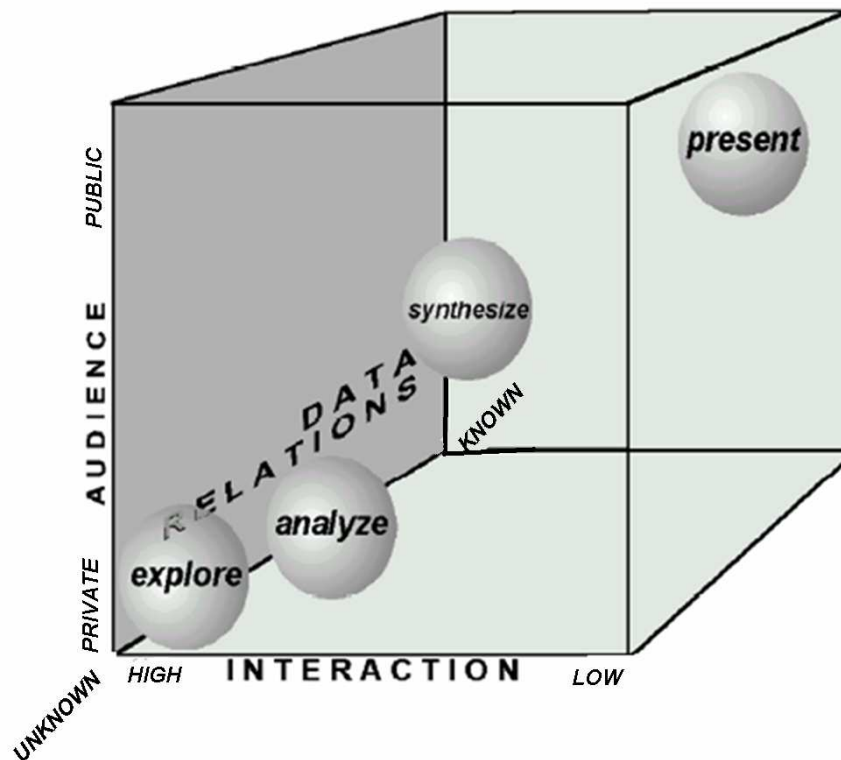


Figure 2.6. Map-use-cube (MacEachren et Kraak, 1997).

Ainsi dans ce contexte les cartes interactives deviennent des outils incontournables pour l'exploration et l'analyse des données géographiques. Une carte est interactive si elle fonctionne comme une interface vers d'autres données. A travers un simple clic sur l'objet géographique, la carte peut renvoyer d'autres informations comme des cartes, des images, documents multimédia, etc. L'interactivité signifie aussi que la carte peut adapter son contenu et son affichage en fonction des actions de l'utilisateur.

Un aspect fondamental de la géovisualisation est la possibilité de représenter les données des différentes façons. Celles-ci ne sont pas affichées seulement avec des techniques classiques, ainsi l'utilisateur peut expérimenter de nouveaux paradigmes de visualisation et d'interaction (par exemple l'usage des landscapes à la place des cartes topographiques, une carte 3D à la place d'une carte choroplèthe). A chaque méthode correspond une vision différente des données et par conséquent un autre aperçu différent du phénomène spatial. Deux exemples de techniques avancées de visualisation et interaction sont montrés en Figure 2.7. La Figure 2.7a montre la technique du « Space-Time Cube » (Kraak et Koussoulakou, 2005) pour visualiser la localisation spatio-temporelle des fouilles archéologiques. La Figure 2.7b montre l'utilisation d'une Matrice multiforme bivariée (Multiform bivariate matrix) (MacEachren, et al., 2004) pour l'étude de la mortalité liée au cancer.

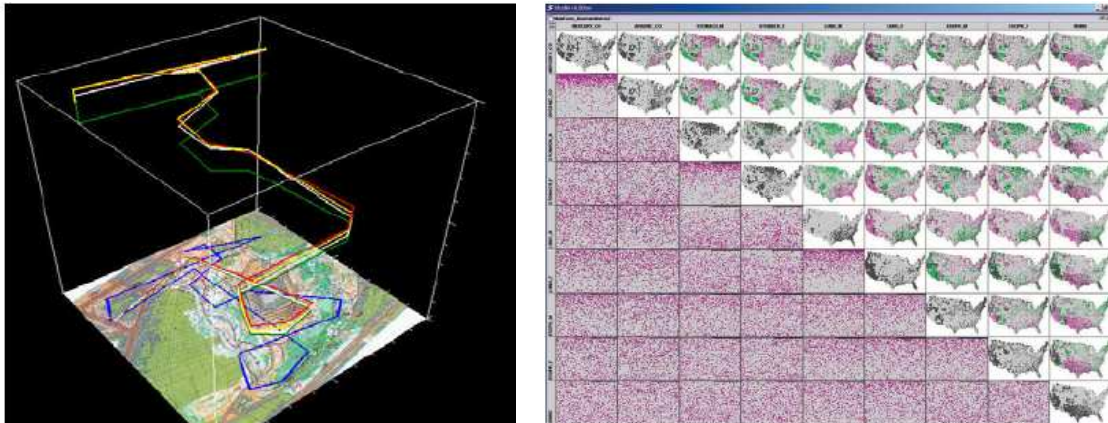


Figure 2.7. a) Space-Time Cube (Kraak et Koussoulakou, 2005) b) Matrice multiforme bivariée (Multiform bivariate matrix) (MacEachren, et al., 2004).

En conclusion, la géovisualisation représente un instrument qui facilite l'extériorisation des connaissances géographiques cachées dans les données géographiques.

### 2.3 Conclusions

Un Système d'Information Géographique permet de mémoriser, analyser et visualiser les données géographiques. La prise en compte de la nature particulière de l'information géographique implique une reformulation des outils et des techniques classiques de mémorisation, d'analyse et de visualisation.

Les aspects spatiaux et sémantiques de l'information géographique, le pouvoir expressif des cartes et le caractère itératif et flexible du processus d'analyse spatiale caractérisent la l'analyse de l'information géographique. Ainsi l'intégration de l'information géographique au sein d'un système d'aide à la décision classique implique une reformulation des caractéristiques principales de ce dernier, en termes de modèles d'analyse et de techniques de visualisation et interaction.

# *Chapitre 3 : L'OLAP Spatial*

Le terme OLAP Spatial (SOLAP) identifie un type de système d'aide à la décision spatiale qui intègre les techniques d'analyse de l'OLAP et des systèmes d'information géographique. Le SOLAP augmente les capacités d'analyse des systèmes OLAP classiques et il implique une reformulation des concepts des entrepôts de données et de l'analyse en ligne d'un point de vue formel et d'implémentation.

Dans ce chapitre, nous décrivons tout d'abord les caractéristiques fondamentales des systèmes d'aide à la décision spatiale, et ensuite nous présentons les concepts principaux des entrepôts de données spatiales et du SOLAP. Nous nous intéressons plus particulièrement aux définitions de dimension et de mesure spatiale. Le chapitre présente aussi un panorama des modèles formels pour les bases de données spatio-multidimensionnelles et des systèmes SOLAP existants.

## **3.1 De l'OLAP à l'OLAP Spatial**

Dans cette section nous introduisons les caractéristiques principales d'un système d'aide à la décision spatiale et les avantages liés à l'introduction de la composante cartographique dans l'OLAP.

### **3.1.1 Les systèmes d'aide à la décision spatiale**

Dans le contexte de l'exploration et d'analyse interactive de données spatiales, qui représentent 80% des données transactionnelles (Franklin, 1992), les cartes ne sont pas seulement de simples outils de visualisation, mais elles sont des instruments actifs qui stimulent l'utilisateur dans son processus mental d'analyse (MacEachren et Kraak, 2001). Les Systèmes d'Aide à la Décision (SAD), et les systèmes OLAP en particulier, ne présentent aucun instrument pour la gestion des données spatiales. Par conséquent, les outils OLAP, en prenant peu en compte la composante spatiale et son pouvoir d'expression et d'analyse, manquent d'un instrument fondamental d'analyse et d'exploration qui peut aider l'utilisateur dans le processus décisionnel (Caron, 1998). A la base des SIG existe une technologie OLTP, contrairement aux systèmes d'entrepôts de données qui s'appuient sur une technologie OLAP (cf chap 1). Les SIG ne peuvent pas être considérés comme de véritables SAD, car même s'ils incluent des fonctionnalités avancées d'analyse, ils manquent d'une interface simple et intuitive pour visualiser et requêter les données, les temps d'analyse sont longs et ils ne sont pas conçus pour fournir une vision agrégée des données (Keenan, 1996). Un nouveau type de SAD a alors été conçu : les Systèmes d'Aide à la Décision Spatiale (SADS).

Un SADS étend les fonctionnalités des systèmes d'information géographique (gestion de la donnée spatiale, affichage cartographique et graphique, outils d'analyse spatiale), en fournissant des modèles d'analyse, des interfaces simples et flexibles, et gérant des structures complexes de données spatiales (Armstrong et Denshman, 1990), (Goodchild et Denshman, 1990). Ils aident les décideurs à résoudre de problèmes et prendre de décisions dans les applications où la composante spatiale joue un rôle fondamental comme le géomarketing, l'environnement, etc. Différents types de SADS ont été développés, comme par exemple les systèmes « Exploratory Spatial Data Analysis » (Andrienko et al., 2003), qui intègrent des techniques d'interaction et de visualisation cartographique, et des méthodes d'analyse spatiales, ou les systèmes « Spatial Data Mining » (Compieta et al., 2007) qui fournissent des techniques de fouille de données spatiales et visualisation cartographique. Chacun s'adresse à des problématiques décisionnelles spatiales particulières. Parmi ces solutions, on trouve aussi le SOLAP qui vise à intégrer la donnée spatiale dans l'OLAP.

### 3.1.2 Les avantages de l'OLAP Spatial

Le Spatial OLAP a été défini par Yvan Bédard comme « *une plateforme visuelle conçue spécialement pour supporter une analyse spatio-temporelle rapide et efficace à travers une approche multidimensionnelle qui comprend des niveaux d'agrégation cartographiques, graphiques et tabulaires* » (Bédard, 1997).

La visualisation des mesures sur une carte permet de comprendre la distribution géographique d'un phénomène qui, souvent, peut être différente de l'espace géographique identifié par la structure hiérarchique définie par la dimension géographique. De plus, l'affichage cartographique révèle des informations spatiales (relations spatiales et informations métriques) qu'une simple étiquette textuelle ou un affichage graphique n'aurait jamais montrées. La visualisation cartographique dans un contexte multidimensionnel permet aussi de mettre en relation les différents phénomènes spatiaux par rapport aux axes d'analyse alphanumériques et de comparer ces phénomènes à diverses granularités géographiques.

La composante cartographique dans l'OLAP représente un instrument de visualisation et surtout d'analyse, qui permet à l'utilisateur de voir et comprendre les données spatio-multidimensionnelles, et elle constitue une interface vers l'entrepôt de données spatiale (Bédard et al., 2005). Utiliser une représentation cartographique des membres pour accéder aux opérations de navigation multidimensionnelle permet à l'utilisateur de se concentrer sur les aspects purement géographiques pendant le processus d'exploration et d'analyse sans qu'il ait besoin d'un modèle mental qui trace dans l'espace géographique les membres des dimensions.

## 3.2 Concepts principaux de l'OLAP Spatial

Stefanovic et al. définissent un entrepôt de données spatiales comme une collection de données spatiales et thématiques, intégrées, non volatiles et historiées pour la prise de décisions spatiales (Stefanovic et al., 2000). Un entrepôt de données spatiales est une reformulation d'un entrepôt conventionnel. Il contient en même temps des données spatiales et alphanumériques et il reformule les concepts classiques de dimension et de

mesure pour prendre en compte la composante spatiale de l'information géographique en définissant les dimensions et les mesures spatiales. Dans cette section nous décrivons les différentes définitions de dimension spatiale et mesure spatiale proposées dans la littérature.

### 3.2.1 Dimension spatiale

Le terme de dimension spatiale désigne l'introduction de l'information spatiale dans une application décisionnelle en tant qu'axe d'analyse. Diverses définitions de dimension spatiale existent dans la littérature.

#### 3.2.1.1 Les approches existantes

(Bédard et al., 2001), (Rivest et al., 2003) définissent une dimension spatiale comme :

- « *non géométrique* » si les membres sont localisés seulement par un label textuel, comme une adresse ou un code postal.
- « *géométrique* » si tous les membres présentent une composante spatiale (point, ligne, polygone, etc.)
- « *mixte* » s'il y a des niveaux avec la composante spatiale et des niveaux purement textuels

Un exemple de dimension spatiale géométrique représentant la subdivision administrative de France (département, région et pays) est montrée en Figure 3.1. Le schéma de la hiérarchie est montré en Figure 3.1a et un exemple des membres avec leur représentation cartographique est représenté en Figure 3.1b.

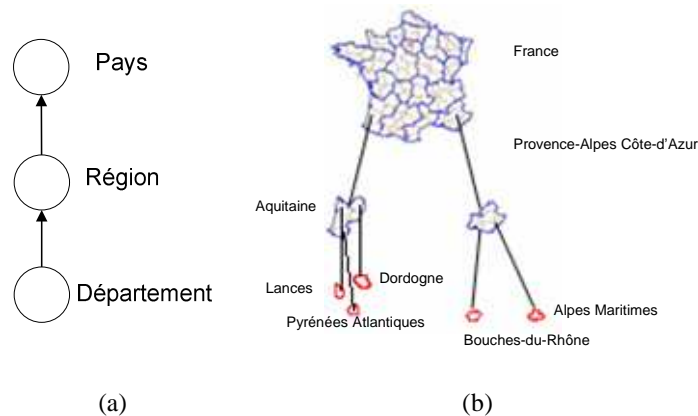


Figure 3.1. Dimension spatiale géométrique a) Schéma b) Représentation cartographique des membres.

Dans (Malinowski et Zimányi, 2005), les auteurs introduisent le concept de dimension spatiale comme un ensemble de hiérarchies spatiales. Une hiérarchie est spatiale s'il y a au moins un niveau qui contient la composante spatiale. Celui-ci est dit niveau spatial. De plus, entre les membres de deux niveaux spatiaux doit exister une relation topologique

d'inclusion ou d'intersection. Une hiérarchie spatiale peut être totalement spatiale si tous les niveaux sont spatiaux, partiellement spatiale s'il y a au moins un niveau non spatial.

De leur côté, Fidalgo et al. (Fidalgo et al. 2004) définissent deux types de dimensions : géographique et hybride. Une dimension géographique contient seulement des niveaux spatiaux. Une dimension est hybride si elle présente à la fois des niveaux spatiaux et alphanumériques. De plus, en se concentrant sur la présence ou non des attributs descriptifs dans un même niveau, ils introduisent une sous-classification de ces dimensions : primitive si seuls les attributs spatiaux sont présents, et composée si des attributs alphanumériques sont aussi inclus.

### **3.2.1.2 Bilan**

(Malinowski et Zimányi, 2005) et (Fidalgo et al. 2004) définissent respectivement une dimension spatiale ou hybride si au moins un niveau contient la composante spatiale. Ils distinguent ces dimensions par rapport au nombre des niveaux qui présentent la géométrie (partiellement spatiale, hybride, totalement spatiale et géographique). (Bédard, et al. 2001) appellent spatiale une dimension dont les membres représentent des objets géographiques même si la composante spatiale n'est pas représentée explicitement.

Ces auteurs se concentrent tous sur la composante spatiale de l'information géographique. Celle-ci permet d'associer aux niveaux des dimensions une composante cartographique et donc d'exploiter le pouvoir expressif et d'analyse des cartes dans l'analyse multidimensionnelle. De plus, la composante spatiale autorise l'utilisation des prédicats spatiaux dans les opérations de coupe et de prendre en compte les relations topologiques pendant les processus d'agrégation (Jensen et al., 2004), (Malinowski et Zimányi, 2005), (Pedersen et Tryfona, 2001). Par contre, ces nouveaux concepts impliquent une importante reformulation des modèles logiques (Bédard et al., 2001), (Fidalgo et al., 2004), (Malinowski et Zimányi, 2007), (Pestana, et al., 2005) et des techniques de visualisation et d'interaction (Rivest et al., 2005).

### **3.2.2 Mesure spatiale**

Parmi les travaux relatifs aux concepts du SOLAP, de nombreux auteurs se sont intéressés au concept de mesure spatiale. On trouve dans la littérature de nombreuses définitions : la mesure spatiale y est parfois vue comme une collection de pointeurs vers des objets spatiaux (Stefanovic et al., 2000), (Rivest et al., 2001), (Malinowski et Zimányi, 2004) et (Sampaio et al., 2006), comme les résultats d'opérateurs métriques ou topologiques spatiaux, par exemple la distance entre deux régions (Rivest et al., 2001), (Malinowski et Zimányi, 2004), ou comme un membre spatial d'une dimension (Marchand et al., 2003). Cette liste montre à quel point le concept de mesure spatiale peut être envisagé de multiples façons : géométrie des objets spatiaux, résultats d'opérateurs spatiaux, et attributs métriques d'un objet spatial. Par voie de conséquence, les fonctions d'agrégation spatiales sont elles aussi très différentes et plus complexes que les fonctions classiques (COUNT, SUM, etc.) utilisées dans les systèmes OLAP. Les systèmes SOLAP proposent d'utiliser par exemple l'union, l'intersection ou le barycentre. Leur utilisation permet d'avoir des descriptions de synthèse du phénomène spatial.

### 3.2.2.1 La mesure spatiale vue comme une liste d'objets spatiaux

Nous présentons l'exemple décrit par (Stefanovic et al., 2000) car il est représentatif de ces types de travaux. En utilisant le schéma de la Figure 3.2, les auteurs se proposent d'étudier les précipitations sur une zone géographique donnée (La Colombie Britannique). Pour chaque relevé de précipitation sur une région (dimension avec une composante géométrique « Localisation »), la température (dimension « Température »), la date (dimension « Temps ») et le niveau de précipitation (dimension « Précipitation ») sont enregistrés. Les mesures « Superficie » et « Nombre de régions » sont de type numérique et représentent respectivement la superficie totale et le nombre des régions examinées. La dernière mesure « région-map » est la mesure spatiale qui représente un pointeur sur une zone d'une province (« Station ») où le relevé de précipitation a été effectuée. « Station » dans la dimension « Localisation », et « region\_map » contiennent des pointeurs vers des objets spatiaux désignant des régions de la même carte. Les auteurs définissent la fonction d'agrégation sur la mesure spatiale de cette façon : si les régions à agréger sont contiguës alors on applique l'opérateur topologique d'union, sinon on maintient une liste de l'ensemble des régions.

Ce modèle multidimensionnel permet d'analyser la localisation des phénomènes météorologiques en fonction du temps, des régions et de la température.

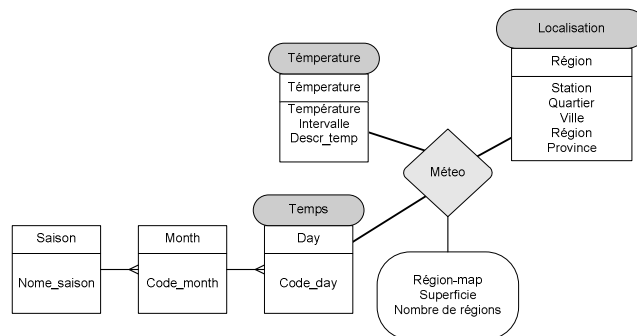


Figure 3.2. Schéma de l'entrepôt « Météo » (Stefanovic et al., 2000).

Dans cette proposition, la mesure est un objet spatial réduit à sa part géométrique, et à des attributs numériques qui en sont directement dérivés comme la superficie et le nombre régions. Pour intégrer aux mesures spatiales leurs caractéristiques alphanumériques, le nom de la région à laquelle la station appartient, etc., Stefanovic et al. répliquent la mesure spatiale dans le niveau le plus détaillé de la dimension spatiale. Les attributs de la dimension représentent alors les informations alphanumériques des mesures spatiales (détaillées ou agrées).

Une autre approche pour intégrer les caractéristiques géométriques à un fait spatial consiste à éliminer la composante spatiale du fait et à la reporter dans une dimension comme présentée en (Fidalgo et al., 2004). La table de faits contient alors un pointeur vers les membres de la dimension spatiale. Les auteurs n'utilisent pas explicitement une mesure spatiale afin d'éliminer la redondance de l'information géométrique présente dans la mesure et dans la dimension spatiale, et surtout pour permettre l'implémentation des

applications spatio-multidimensionnelles à travers les serveurs OLAP classiques. De la même façon (Marchand et al., 2004) définissent une mesure spatiale comme un membre d'une dimension spatiale.

(Rivest et al., 2001), (Malinowski et Zimányi, 2004) et (Sampaio et al., 2006) définissent une mesure spatiale comme un objet géométrique qui est un attribut spatial du fait, et qui, contrairement aux modèles de (Fidalgo et al., 2004), (Stefanovic et al., 2000) et (Marchand et al., 2003), peut ne pas être répliqué dans une dimension spatiale. Un exemple est montré en Figure 3.3. L'application spatio-multidimensionnelle concerne les accidents routiers. Elle permet d'analyser la localisation des accidents en fonction du temps, du type d'assurance et des clients.

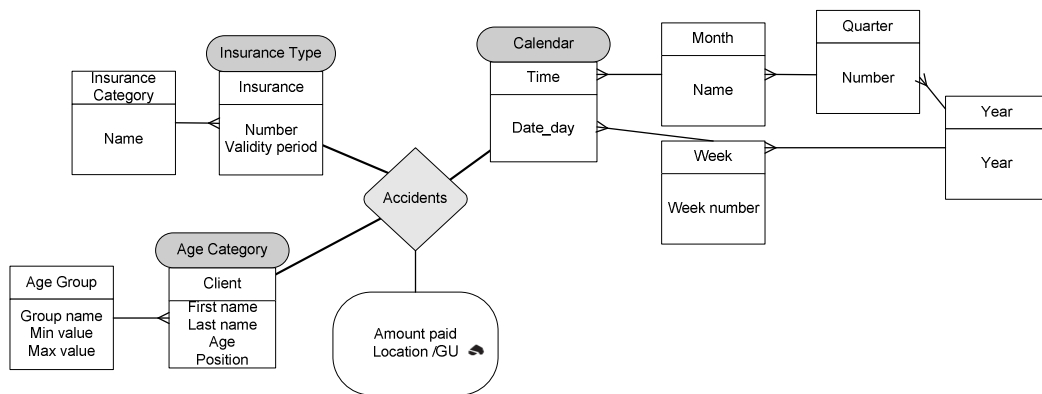


Figure 3.3. Application spatio-multidimensionnelle avec de dimensions alphanumériques et une mesure spatiale (Malinowski et Zimányi, 2004)

Elle présente trois dimensions : le temps (« Calendar »), l'âge du client (« Age Category ») et le type d'assurance (« Insurance Type »). La mesure spatiale « Location », agrégée avec l'agrégation spatiale d'union topologique (« GU »), représente le lieu de l'accident. Une requête possible est : « Pour chaque jour à quel endroit ont eu lieu les accidents de chaque client ? ».

### 3.2.2.2 La mesure spatiale vue comme le résultat des opérateurs spatiaux

(Rivest et al., 2001) et (Malinowski et Zimányi, 2004) définissent aussi la mesure spatiale comme le résultat d'opérations spatiales appliquées aux membres des dimensions spatiales, par exemple la distance entre deux objets spatiaux. Cette définition n'exclut pas la définition de mesure spatiale comme définie au paragraphe précédent.

Un exemple d'application SOLAP basée sur ce concept, qui concerne les coûts de réparation des autoroutes (Malinowski et Zimányi, 2004) est montré en Figure 3.4.



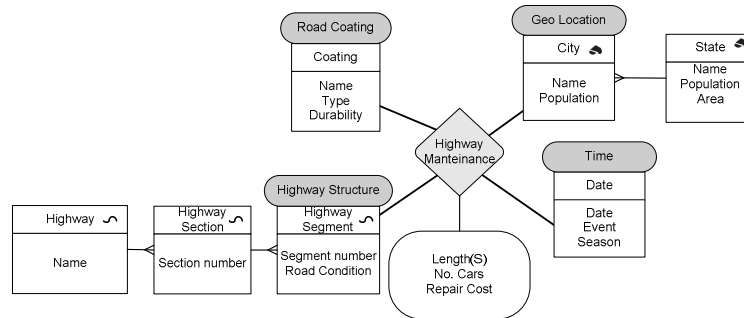


Figure 3.4. Schéma conceptuel d'un entrepôt de données avec une mesure résultat d'une opération spatiale (Malinowski et Zimányi, 2004)

Cette application présente une table de faits « Highways Maintenance » avec trois mesures : le coût de réparation, la longueur de la portion d'autoroute qui appartient à une ville « Length(S) » (mesure spatiale), et le nombre de voitures. Ces mesures sont agrégées en utilisant la somme. Les dimensions qui représentent la composante géométrique sont la localisation « Geo Location » et les autoroutes « Highways structure ». Les deux autres dimensions classiques sont le temps « Time » et le type de revêtement « Road Coating ». Cette application multidimensionnelle permet d'analyser les coûts de réparation des parties d'autoroutes en fonction du temps, des autoroutes et des villes. Deux requêtes possibles sont alors : « *Quelle est la longueur (mesure spatiale) des parties des autoroutes qui passent par des villes ?* » et « *Quel est le coût de réparation de toutes les autoroutes (mesure numérique) pour le chaque année ?* ».

Dans ce type d'application spatio-multidimensionnelle, la table de faits définit une jointure spatiale entre les dimensions spatiales. Cette table de faits permet d'avoir une vision multidimensionnelle des relations spatiales. Elle autorise le calcul des mesures spatiales pour les membres de tous les niveaux des hiérarchies des dimensions, comme par exemple les parties des autoroutes qui passent par des villes ou les parties des autoroutes qui passent par des pays.

### 3.2.2.3 Bilan

Le fait que de nombreux scénarii puissent être représentés grâce aux mesures spatiales montre l'intérêt de ce nouveau concept. Les mesures spatiales permettent de faire de l'information spatiale le sujet de l'analyse multidimensionnelle. Les auteurs identifient deux types de mesure spatiale : une collection d'objets spatiaux et/ou le résultat d'une opération spatiale.

Pour modéliser la mesure spatiale comme une collection de géométries il existe différentes possibilités. (Stefanovic et al., 2000) répliquent la mesure spatiale dans une dimension spatiale, (Fidalgo et al., 2004) et (Marchand et al., 2003) représentent la mesure spatiale comme un attribut d'un objet géographique qui est un membre de dimension, et (Rivest et al., 2001) (Malinowski et Zimányi, 2004) et (Sampaio et al., 2006) modélisent la mesure spatiale comme un attribut spatial de la relation représentant les faits.

(Rivest et al., 2001), (Malinowski et Zimányi, 2004) définissent une mesure spatiale aussi comme le résultat d'une opération spatiale appliquée aux membres des dimensions et ils la représentent comme un attribut de la relation spatiale qui compose la relation représentant le fait.

De la même manière que pour les dimensions spatiales, les auteurs utilisent dans une application spatio-multidimensionnelle seulement la composante spatiale de l'information géographique en tant que sujet d'analyse.

### 3.2.3 Opérateurs de navigation spatio-multidimensionnelle

L'introduction des données spatiales dans les dimensions d'entrepôts de données spatiales a mène différents auteurs à la définition d'opérateurs d'analyse spatio-multidimensionnelles.

(Rivest et al., 2005), (Sampaio, et al., 2006), (Matias et Moura-Pires, 2007), (Scotch et Parmanto, 2005), (Hernandez et al., 2005) appellent « spatial-drill down » et « spatial roll-up » les opérateurs de forage sur une dimension qui présente des niveaux avec une composante spatiale. (Rivest et al., 2005) définissent une opération de « spatial drill-across » comme une opération de coupe sur un ensemble de membres de la dimension spatiale, sélectionnés directement par l'utilisateur à travers l'interface visuelle du système SOLAP. Le terme « spatial » indique simplement le fait que l'opération de navigation multidimensionnelle (forage ou coupe) s'applique à une dimension avec des attributs géométriques.

En plus des opérateurs multidimensionnels « classiques », quelques auteurs ont défini des opérateurs multidimensionnels qui utilisent explicitement les relations spatiales. (Matias et Moura-Pires, 2007) appellent « spatial slice » une opération de coupe sur un hypercube spatial. Cette opération peut porter sur une dimension classique comme la dimension temporelle, ou sur la dimension qui contient l'information spatiale (dimension spatiale). Dans le deuxième cas, un prédicat spatial ou alphanumérique portant sur objets géographiques de la dimension peut être appliqué. (Scotch et Parmanto, 2005) introduisent respectivement deux nouveaux opérateurs de coupe : le « buffer » et le « spatial drill-out ». Le « buffer » utilise l'opérateur d'analyse spatiale de buffer, qui crée une zone tampon autour d'un membre, pour sélectionner des membres de la dimension spatiale. Le « spatial drill-out » sélectionne tous les membres adjacents au membre sur lequel cet opérateur est appliqué. (Sampaio, et al., 2006) et (Colonnese et al., 2005) appellent « spatial slice » une opération de coupe qui utilise un prédicat spatial.

## 3.3 Un panorama des modèles conceptuels pour les entrepôts de données spatiales

Dans cette section nous présentons les différents modèles formels pour les bases de données spatio-multidimensionnelles proposés en littérature.

### 3.3.1 Pourrabas

(Pourrabas, 2003) présente un modèle formel qui intègre une base de données spatiales objets et une base de données multidimensionnelles. Cette solution permet de répondre aux requêtes qui portent sur des données stockées dans une base de données multidimensionnelle et une base de données spatiales, de façon transparente à l'utilisateur.

Pour effectuer ce lien, l'auteur utilise la dimension spatiale, et il étend les structures de données géographiques avec des attributs particuliers appelés attributs fonctionnels. Les membres de la dimension spatiale sont contraints par une relation d'inclusion complète où les géométries des membres des niveaux moins détaillés sont les résultats de la fusion des géométries de leurs descendants.

### 3.3.2 Malinowsky et Zimányi

Le travail présenté dans (Malinowsky et Zimányi, 2004), (Malinowsky et Zimányi, 2005) et (Malinowsky, 2006) étend le modèle multidimensionnel MultiDim (Malinowsky et Zimányi, 2004b) avec les concepts de dimension, hiérarchie et mesure spatiales. Le modèle multidimensionnel est défini comme un ensemble fini de dimensions et une relation de fait (« fact relationships »). Une dimension est composée par plusieurs hiérarchies qui sont constituées par un ou différents niveaux représentés par des entités. Un « fact relationship » représente le sujet de l'analyse et il modélise une relation plusieurs à plusieurs entre les différents niveaux les plus détaillés des dimensions. Chaque fait est décrit par un ensemble de mesures auxquelles sont associées des fonctions d'agrégation. Une hiérarchie est dite spatiale si au moins un niveau contient un attribut géométrique. Deux niveaux spatiaux sont liés par une relation topologique d'inclusion ou d'intersection. Pour ce qui concerne les mesures, on distingue les mesures numériques de mesures spatiales. Une mesure spatiale est représentée par une géométrie ou par le résultat d'une fonction spatiale. A chaque élément du modèle est associée une représentation graphique et/ou iconique, comme montré en Figure 3.5.

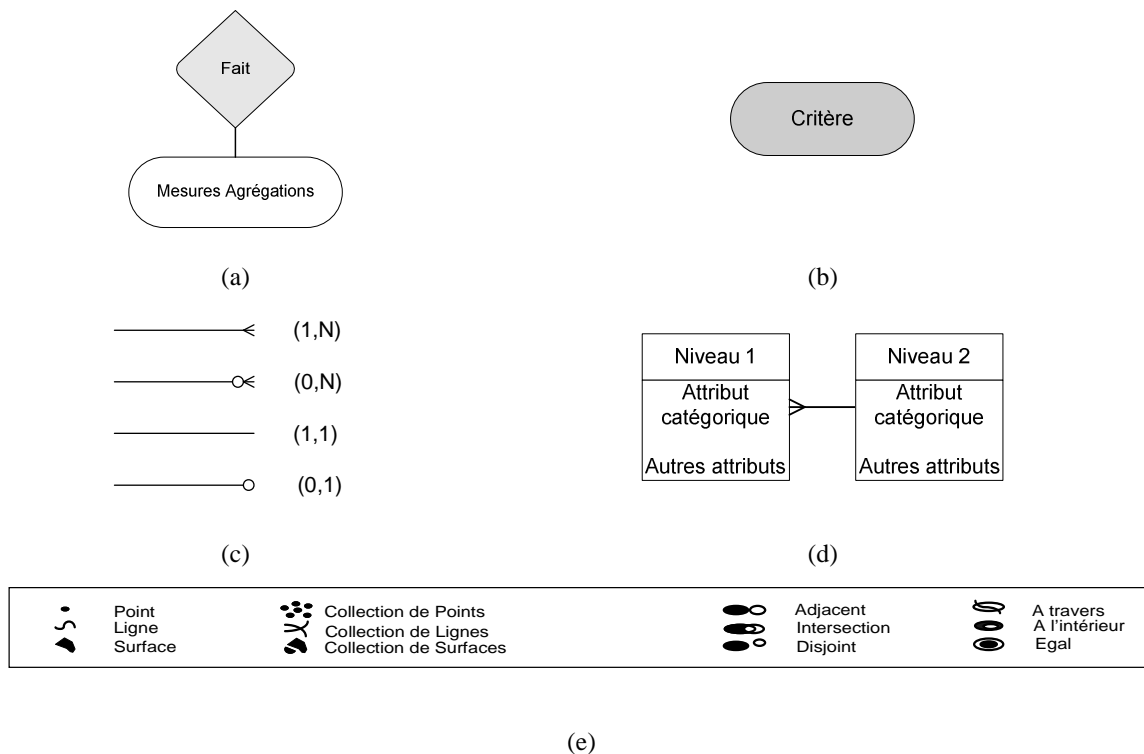


Figure 3.5. Représentation graphique du modèle pour entrepôts de données spatiales présenté dans (Malinowsky et Zimányi, 2004) a) Fait et Mesures b) Dimension c) Cardinalités d) Niveaux de dimension e) Icônes spatiales.

A notre connaissance le modèle représente le seul modèle existant qui fournit un formalisme graphique pour la représentation des données et des relations spatiales dans un modèle multidimensionnel. Par contre, le modèle ne permet pas de représenter les mesures dérivées car aucun langage de calcul n'est présent dans le modèle, contrairement, par exemple, aux modèles basés sur UML qui utilisent OCL (Abello et al. 2006) et (Trujillo et al., 2001). Les mesures sont représentées comme les attributs d'une relation plusieurs à plusieurs entre les dimensions. Cela implique que les relations plusieurs à plusieurs entre faits et dimensions ne peuvent pas être modélisées, car chaque combinaison de membres de dimensions identifié de façon unique un fait.

### 3.3.3 Damiani et Spaccapietra

(Damiani et Spaccapietra, 2006) présentent le modèle MuSD (Multigranular Spatial Data warehouse) et une algèbre pour modéliser et requêter les entrepôts de données spatiales, étendant le modèle présenté dans (Vassiliadis, 1998). MuSD est défini en accord avec les standards de la modélisation géographique et il enrichit les concepts des entrepôts de données classiques avec les concepts de dimension et mesure spatiale. L'innovation de ce modèle est la représentation de la mesure spatiale à plusieurs granularités géométriques.

Le modèle définit une hiérarchie comme un treillis de niveaux représenté par un ordre partiel  $\leq_l v$  où deux niveaux  $Lv1$  et  $Lv2$  sont  $Lv1 \leq_l v Lv2$  ( $Lv1$  est plus détaillé que  $Lv2$ ) si pour chaque membre de  $Lv1$ , il existe un membre de  $Lv2$  tel que  $v1 \leq_l v v2$ . En accord avec

le standard OGC, un niveau est dit spatial s'il représente un ensemble d'objets avec des attributs alphanumériques et un attribut spatial. Une mesure spatiale est définie comme une hiérarchie où tous les niveaux sont spatiaux. Les auteurs introduisent le concept de « Multigranular spatial schema ». Il s'agit d'un ensemble de dimensions, de mesures classiques et d'une mesure spatiale. Le « Multigranular spatial schema » représente un ensemble d'hypercubes. Un hypercube est défini pour chaque niveau de la mesure spatiale. Pour exploiter l'appartenance de la mesure spatiale à une hiérarchie, l'algèbre fournit l'opérateur « Measure Climbing » qui permet de monter dans la hiérarchie de la mesure spatiale et donc de naviguer entre les différents hypercubes.

### 3.3.4 Jensen et al.

Un modèle et une algèbre pour les applications multidimensionnelles pour les services basés sur la localisation, sont décrits dans (Jensen, et al., 2004). Ce travail étend le modèle pour les données classiques présenté dans (Pedersen, et al., 2001). La particularité de cette approche est la prise en compte des relations topologiques d'inclusion partielle ou totale entre les membres d'une dimension spatiale. Le modèle définit un « n-dimensional fact schema » comme deux tuples représentant la table de faits et un ensemble de schémas de dimensions (« Dimension Type »). Une dimension est un ensemble de niveaux appelés « catégories », et deux relations d'ordre partiel. Ces deux relations sont utilisées pour modéliser respectivement les inclusions totales et partielles. Une instance d'une dimension est un ordre total sur les membres des « catégories ». De plus, pour tous les membres, il existe une fonction qui associe à chaque couple des membres ( $e_1$ ,  $e_2$ ) une valeur numérique entre 0 et 1 qui représente le degré d'inclusion : 1 si  $e_1$  est totalement contenu en  $e_2$ , 0 si  $e_1$  peut être partiellement contenu en  $e_2$ , et entre 0 et 1 si  $e_1$  est partiellement contenu en  $e_2$ . Le modèle fournit aussi un algèbre avec des opérateurs de sélection, union, et agrégation qui prennent en compte la relation d'inclusion partielle. Plus précisément, les prédicats de sélection dépendent du degré d'inclusion, l'union recalcule cette valeur pour les nouveaux membres et l'opération d'agrégation peut, grâce à cette valeur, établir plusieurs stratégies de regroupement.

Ce modèle introduit le concept d'imprécision dans les hiérarchies des dimensions, permettant de représenter les relations topologiques d'inclusion et d'intersection entre les membres des dimensions spatiales, même s'il n'utilise pas du modèle particulier pour les données spatiales.

### 3.3.5 Ahmed et Miquel

Contrairement à tous les modèles de SOLAP qui utilisent le modèle vectoriel pour la représentation de l'information géographique, le travail d'Ahmed et Miquel (Ahmed, et Miquel, 2005), introduit une vision continue de l'espace géographique (cf chap 2 § 2.1) dans les entrepôts de données spatiales. Les auteurs étendent le modèle présenté dans (Vassiliadis, 1998). Le modèle est basé sur la notion de « cube de base discret » qui est un tuple  $\langle Db, Lb, Rb \rangle$  où le  $Db$  est une liste de dimensions comprenant une mesure  $M$ ,  $Lb$  est la liste des niveaux les plus détaillés des dimensions et  $Rb$  est un ensemble de cellules représentées comme l'ensemble des tuples contenant les membres des niveaux des dimensions  $x_i$  et les mesures  $m_i$  sous la forme  $x = [x_1, x_2 \dots x_n, m]$ . Afin de créer une

représentation continue du cube de base, des valeurs estimées sont dérivées du cube de base discret. Les mesures estimées sont calculées grâce à des fonctions d'interpolation.

Grâce à cette vision continue de l'espace, le modèle permet de gérer les données manquantes et erronées dans les bases de données multidimensionnelles spatiales et d'avoir des mesures continues sur l'espace.

### 3.3.6 Bilan

Dans cette section nous avons décrit les modèles formels SOLAP. (Pourrabas, 2003) modélise la coopération entre une base de données géographiques et une base de données multidimensionnelle à travers une extension des structures de données géographiques. (Jensen, et al., 2004) développent un modèle et une algèbre pour gérer l'imprécision liée aux relations topologiques d'intersection entre les membres. (Damiani et Spaccapietra, 2006) fournissent un modèle et une algèbre qui permet de naviguer entre plusieurs hypercubes qui présentent la mesure spatiale à différentes granularités spatiales. (Malinowsky et Zimányi, 2004) définissent un modèle formel, avec une représentation graphique pour les entrepôts de données spatiales. (Ahmed et Miquel, 2005) se concentrent sur une vision continue de l'information géographique contrairement à toutes les autres approches. Ils intègrent des fonctions d'interpolation à un modèle et une algèbre multidimensionnelle classiques.

Tous ces modèles se concentrent sur la modélisation des concepts de dimension et de mesure spatiale et ils se focalisent sur des aspects particuliers des données géographiques comme les relations topologiques, la continuité et la multigranularité spatiale. En revanche, ils réduisent l'information géographique à sa composante spatiale et ils sous-emploient la composante sémantique. En d'autres termes, la prise en compte des divers attributs des objets géographiques : spatial et descriptifs, et les relations issues des opérateurs de généralisation de l'information géographique, restent de problématiques ouvertes.

## 3.4 Les outils OLAP Spatial

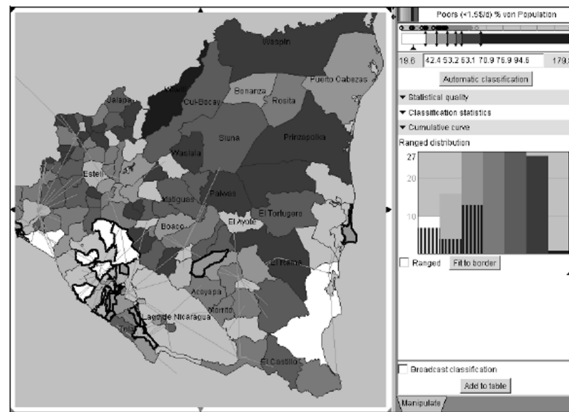
Un outil SOLAP repose sur l'intégration des fonctionnalités SIG et OLAP (Kouba, et al., 2000), (Rivest et al., 2005). La composante cartographique est utilisée pour visualiser les membres de dimensions et/ou les mesures avec une composante spatiale, pour représenter les mesures alphanumériques, grâce à des cartes thématiques, et pour accéder aux opérations de navigation multidimensionnelle. Différents systèmes SOLAP, qui peuvent être classifiés en trois différentes typologies, ont été développés.

### 3.4.1 Typologies

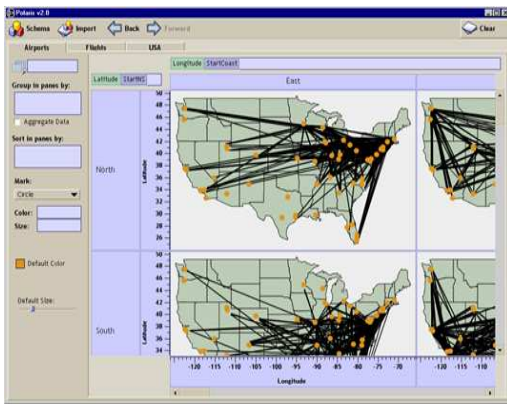
En (Rivest, 2000) et (Bédard, et al., 2005) les solutions SOLAP sont regroupées en trois grandes classes: Olap dominant, SIG dominant et OLAP-SIG intégrée.

#### 3.4.1.1 SIG dominant

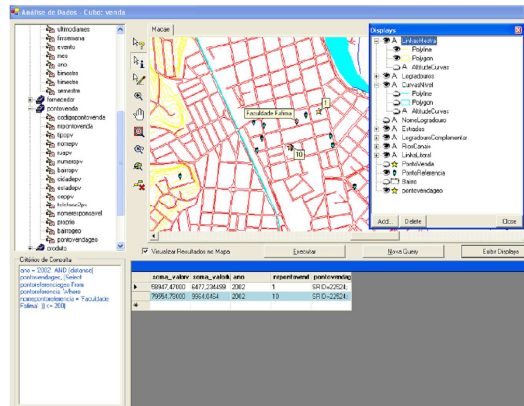
Dans les solutions SIG dominantes comme (Hernandez et al., 2005) (Figure 3.6a), le serveur OLAP est simulé grâce à une base de données relationnelle modélisée sous forme d'étoile.



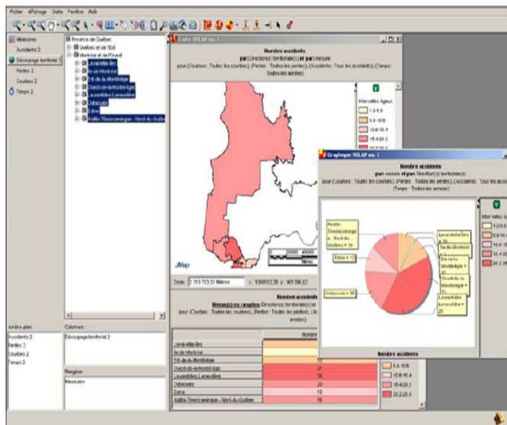
(a)



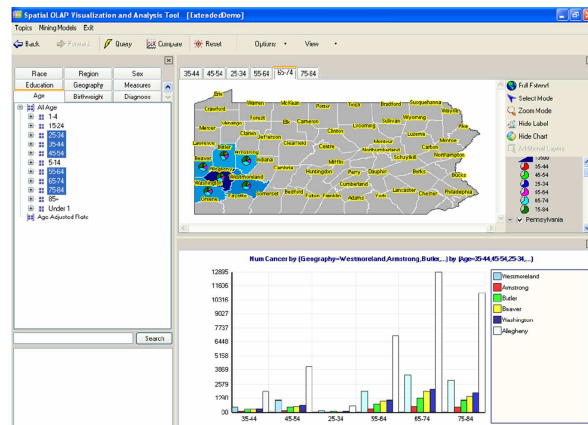
(b)



(c)



(d)



(e)

Figure 3.6. Interfaces des systèmes SOLAP a) CommonGIS (Hernández et al., 2005) b) Polaris (Stolte et al., 2002) c) PostGeoLAP (Colonnese et al., 2005) d) JMap (Jmap, 2007) e) SOVAT (Scotch et Parmanto, 2006).

Les solutions SIG dominantes, comme décrit par (Bédard et al., 2005), offrent toutes les fonctionnalités d'un outil SIG : stockage, analyse et visualisation des données spatiales. Par contre, elles doivent inclure, dans la base de données, des éléments permettant d'implémenter les opérations OLAP de forage et de coupe, puisqu'il n'existe pas de serveur OLAP pour gérer ces opérations. De plus, toutes les fonctionnalités avancées OLAP comme l'utilisation de mesures dérivées, ne sont pas présentes dans ce type d'outil, ce qui limite ses capacités d'analyse multidimensionnelles.

### **3.4.1.2 OLAP dominant**

Les outils OLAP dominants comme (Stolte et al., 2003) (Figure 3.6b) utilisent un système OLAP et offrent toutes les fonctionnalités classiques pour l'analyse multidimensionnelle. Par contre, les fonctionnalités SIG sont limitées à une simple représentation cartographique des mesures et des dimensions spatiales, à la navigation cartographique (i.e. déplacement et changement de couche) et à la sélection d'objets géographiques (Bédard et al., 2005). Ces solutions ne présentent aucun instrument pour l'analyse spatiale ou d'autres fonctionnalités avancées SIG. Ceux-ci, comme montré par plusieurs travaux (Hernández et al., 2005), (Bédard et al., 2002) sont nécessaires et complémentaires à l'analyse spatio-multidimensionnelle. De plus, dans ces solutions les opérateurs de forage sur la dimension spatiale sont inexistantes ou limités. Il n'est pas possible par exemple, d'obtenir une carte avec des membres spatiaux de différents niveaux, grâce à une simple opération de forage sur un seul membre spatial (Bédard et al., 2005).

Par conséquent, dans les solutions OLAP dominant, l'information géographique n'est pas exploitée, ce qui implique que ce type de solution présente quelques limitations pour une analyse spatio-multidimensionnelle concrète. L'utilisateur ne peut pas comparer visuellement les mesures associées aux membres de différents niveaux spatiaux, ou encore il n'a pas accès aux outils SIG pour analyser les données géographiques.

### **3.4.1.3 OLAP-SIG intégrée**

Les solutions OLAP-SIG intégrées comme JMap (JMap, 2007) (Figure 3.6d) fusionnent toutes les fonctionnalités des deux différents systèmes dans un seul environnement.

Comme montré dans (Hernández et al., 2005) les fonctionnalités SIG d'analyse et de visualisation sont nécessaires pour l'analyse spatio-multidimensionnelle et elles complètent les fonctionnalités purement OLAP. Les solutions OLAP-SIG intégrées sont alors les plus adaptées pour une analyse spatio-multidimensionnelle réelle et efficace.

Cette intégration peut être vue comme une reformulation des trois niveaux d'une architecture OLAP classique, en utilisant et/ou en ajoutant des fonctionnalités SIG.

Le premier niveau est un entrepôt de données spatiales, qui doit permettre de modéliser les complexes structures de données associées aux dimensions et aux mesures spatiales.

Le deuxième niveau est un serveur OLAP capable de gérer des requêtes spatio-multidimensionnelles. Dans le meilleur des cas, il devrait :



- gérer les hiérarchies complexes (i.e. non-couvrantes, non-strictes, et non-onto) et multiples (Hernández, 2006) et traiter symétriquement mesures et dimensions (Hernández, 2006),
- permettre la définition des mesures spatiales comme objets spatiaux et résultats d'opérations spatiales, et de mesures dérivées à partir des membres et/ou mesures classiques et/ou spatiales,
- considérer les relations spatiales entre les membres de dimensions et éviter de prendre en compte plusieurs fois la même mesure (Malinowsky et Zimányi, 2007a) pendant le processus d'agrégation,
- prendre en compte la composante spatiale des données géographiques lors d'une opération de forage ou de coupe,
- fournir un mécanisme pour calculer les vues matérialisées (Colonnese et al., 2005).

Les fonctionnalités du client OLAP sont réétudiées pour permettre en même temps une analyse spatiale et multidimensionnelle. La composante géométrique doit être visualisée à travers des cartes couplées et synchronisées avec les outils d'affichage typiques de l'OLAP : la table de pivot et les affichages graphiques. En d'autres termes, le résultat d'une requête multidimensionnel doit être représenté sous forme de tableau, de graphiques et à travers la composante cartographique. Celle-là montre, dans le cas d'une dimension géographique, les membres des dimensions sur lesquels sont affichées les mesures numériques, grâce à des affichages graphiques différents, (i.e., bars, camemberts, etc.), ou une carte par mesures, etc. La possibilité d'avoir différentes visualisations cartographiques en même temps, permet d'obtenir diverses visions alternatives des données et par conséquent de faciliter le processus d'analyse (Persson, 2005). Un client SOLAP doit permettre d'explorer les données multidimensionnelles grâce à la simple interaction avec la composante cartographique. Les cartes, dans un outil SOLAP, doivent rendre « visible » l'information et permettre d'explorer et analyser la base de données spatio-multidimensionnelle sans connaître aucun langage particulier de requêtes et/ou la structuration de la base de données spatiale. Par conséquent, la composante cartographique dans le SOLAP, comme montré dans le paragraphe 2.2.4.2, doit être composée par des cartes interactives où la simple interaction génère des actions d'exploration et la visualisation s'adapte aux manipulations de l'utilisateur. De plus, les cartes interactives doivent être couplées avec de techniques de géovisualisation, comme le « MultiMaps » qui est une représentation de plusieurs cartes thématiques sur divers panneaux.

Les trois composantes doivent être synchronisées (Rivest et al., 2001) (Rivest et al., 2005), ou en d'autres termes une action (par exemple un drill-down ou un roll-up) sur une composante (par exemple une carte) doit être propagée aux autres composantes (table de pivot et graphiques). Cette technique de visualisation et d'exploration, connue comme « Multiple coordinate windows » (Sheidermann, 1998) est indispensable au processus d'exploration spatio-temporel (Andrienko et Andrienko, 2003). Chaque diagramme « montre » une caractéristique différente de l'information spatio-multidimensionnelle, par

exemple la table de pivot permet de voir les détails des mesures, les affichages graphiques de les résumer visuellement et les cartes de positionner ces informations dans l'espace.

Enfin, parallèlement aux opérateurs d'exploration, un outil SOLAP doit aussi fournir des fonctionnalités SIG de base et d'analyse spatiale. Ces dernières permettent d'analyser les dimensions et/ou les mesures spatiales grâce aux outils d'analyse et de visualisation. En fait, l'utilisateur peut naviguer dans l'hypercube spatial et en même temps ajouter des cartes en format matriciel ou vectoriel (Rivest et al., 2001), personnaliser leurs affichages, appliquer des opérateurs métriques, requêter la base de données spatiales, appliquer les opérateurs d'analyse spatiale, ou de fouille de données (Scotch et Parmanto, 2006), etc.

### 3.4.2 Un panorama des outils SOLAP existants

Différents systèmes pour intégrer l'information spatiale dans un contexte multidimensionnel ont été développés. Dans cette section nous allons parcourir ces solutions selon le type d'approche utilisée.

#### 3.4.2.1 Solutions SIG dominantes

A notre connaissance la seule solution orientée SIG est présentée dans (Voss et al., 2004) (Hernandez et al., 2005). CommonGis est un logiciel de géovisualisation étendu pour supporter les bases de données multidimensionnelles (Figure 3.6a). Il fournit des fonctionnalités d'analyse multicritère, d'analyse spatiale et des techniques visuelles d'analyse spatio-temporelle en utilisant les séries temporelles. CommonGis a été adapté pour l'analyse des données multidimensionnelles spatiales, où l'information spatiale est incluse en tant qu'axe d'analyse. L'interface, flexible et interactive, ne présente pas la composante tabulaire. En revanche, elle possède une vaste collection de techniques avancées de géovisualisation et autorise les opérations SOLAP de forage et de coupe.

#### 3.4.2.2 Solutions OLAP dominantes

Les solutions SOLAP dominantes peuvent être regroupées en deux grandes classes : les outils qui utilisent des cartes statistiques et les outils qui adoptent des cartes interactives.

La visualisation de données multidimensionnelles à travers des cartes statistiques est à la base de nombreux travaux comme nous décrivons ci-dessous. Le système Polaris (Stolte et al., 2002), (Stolte et al., 2003) (Figure 3.6b) permet la visualisation des mesures alphanumériques à travers des cartes thématiques non interactives, englobées dans les cellules de la table de pivot. PostGeOLAP (Colonnese et al., 2005) (Figure 3.6c) est un système code source libre SOLAP qui gère les mesures numériques et les dimensions spatiales. PostGeOLAP fournit des méthodes pour créer la base de données spatiales multidimensionnelles et les vues matérialisées. Pentaho (Pentaho) et (Mohraz, 2000) (où le système OLAP SAP's Business Information Warehouse est intégré avec un outil de visualisation cartographique), utilisent la composante cartographique seulement pour la visualisation des données géographiques. Enfin, OLAP for ArcGIS (ESRI, 2007) permet de connecter le SIG ArcGIS à de nombreux serveurs OLAP (SAP BW, SAS OLAP Server, and Microsoft SQL Server Analysis Services). OLAP for ArcGIS permet de « voir » les mesures numériques sous forme des graphiques associées aux cartes représentant la dimension spatiale et aussi à travers de tables de pivot. De la même façon qu'OLAP for

ArcGIS, Location Intelligence Component de MapInfo (MapInfoLI, 2007) permet un affichage web des hypercubes définis avec les serveurs OLAP de MicroStrategy, Cognos et Business Object. Nous considérons ces deux dernières solutions comme étant OLAP dominantes car, même si elles combinent fonctionnalités SIG et OLAP, à notre connaissance elles présentent seulement des cartes statistiques. Plusieurs solutions qui utilisent les cartes interactives ont été développées, dont (Silva et al., 2006), (Galigeo, 2007), (Visualiser, 2007) et (Sampaio et al., 2006). (Silva et al., 2006) présentent un outil web SOLAP où la caractéristique principale est l'usage des services web géographiques pour la définition de GeoMDQL. Il s'agit d'un nouveau langage de requêtes qui étend le langage OLAP de Microsoft MDX (MDX, 2007) pour les entrepôts de données spatiales. Le prototype est basé sur le serveur OLAP Mondrian qui a été modifié pour gérer les requêtes GeoMDQL et le client OLAP JPivot couplé avec une simple carte interactive, utilisée pour la représentation des dimensions spatiales. Puisque l'outil utilise la modélisation logique présentée en (Fidalgo et al., 2004), il ne gère pas les mesures spatiales. Dans (Sampaio et al., 2006), les auteurs décrivent un système web SOLAP qui permet d'interroger les entrepôts de données spatiales avec des opérateurs de forage et de sélection sur la dimension spatiale et gérer les mesures spatiales comme un ensemble d'objets spatiaux. Dans cette solution, l'interface web est composée seulement par une composante cartographique, un navigateur pour sélectionner les membres de dimensions, et une zone de texte pour éditer les requêtes multidimensionnelles. Enfin, Visualiser de Cognos (Visualiser, 2007) et Webigeo (Galigeo, 2007) (qui combine le SIG ArcGIS et le serveur OLAP de Business Object), même si présentent des fonctionnalités SIG et OLAP, ne permettent pas d'appliquer le drill-down sur un seul membre de la dimension spatiale, réduisant les possibilités d'interaction fournies par leur composante cartographique.

Enfin, (Shekhar et al., 2001) et (Han et al., 1997) se sont intéressés à l'introduction des techniques de fouilles dans un outil OLAP, et (Pourabbas et Rafanelli, 2002) à la utilisation de langages visuels. En particulier, (Shekhar et al., 2001) développent le « cube operator » étendant le concept d'agrégation aux données spatiales. Leur prototype web, basé sur cet opérateur, est spécialement conçu pour l'observation et la découverte des modèles spatiaux et tendances temporelles dans les entrepôts de données spatiales. Il fournit des méthodes de fouille de données spatiales ainsi que des techniques de visualisation avancées comme le MultiMaps. GeoMiner (Han et al., 1997) est une autre solution de fouille de données spatiales qui utilise une composante SOLAP. Enfin, dans (Pourabbas et Rafanelli, 2002) un langage visuel est utilisé pour interroger une base de données spatio-multidimensionnelle. Dans cette approche, l'utilisateur à l'aide d'un langage iconique formule les opérations de coupe sur les membres de la dimension spatiale.

### ***3.4.2.3 Solutions OLAP-SIG intégrées***

Dans (Rivest et al., 2005) les auteurs décrivent un outil SOLAP qui fournit nombreux types de visualisation cartographique tabulaire et graphique accessibles par plusieurs fenêtres et des fonctionnalités SIG et OLAP. Les composantes visuelles sont synchronisées pour former une interface unique, flexible et interactive. L'outil gère les mesures classiques et les mesures spatiales définies comme l'ensemble de toutes les géométries représentant les objets spatiaux correspondant à une combinaison particulière des membres spatiaux. Cet outil est commercialisé sous le nom de JMap Spatial OLAP (JMap, 2007) (Figure 3.6d). De

nouvelles fonctionnalités ont été récemment introduites, comme par exemple l'association des documents hypermédia aux éléments multidimensionnels spatiaux et/ou non (Bédard, et al. 2006), ou en d'autres termes des images, des photos, des documents textuels, etc. peuvent être associé aux membres des dimensions.

SOVAT (Scotch et Parmanto, 2006) (Figure 3.6e) est un outil OLAP-SIG intégré qui combine des fonctionnalités OLAP, SIG et de fouille de données classiques et spatiales comme par exemple le clustering. De plus, il fournit des méthodes de statistiques et d'analyse spatiale, dont le buffer qui sélectionne un ensemble de membres spatiaux pour couper l'hypercube. L'outil synchronise en ensemble de cartes interactives, et d'affichages graphiques et tabulaires. SOVAT gère les dimensions spatiales et les mesures spatiales comme de collections de géométries.

Le prototype SOLAP présenté dans (Matias et Moura-Pires, 2005), (Matias et Moura-Pires, 2007) se distingue par la possibilité de couper l'hypercube en utilisant des relations spatiales et par la possibilité de superposer deux hypercubes en utilisant leurs dimensions spatiales. Cet outil intègre de fonctionnalités SIG de base, comme la possibilité d'ajouter de couches vectorielles ou matricielles, aux fonctionnalités purement OLAP. L'interface combine et synchronise des fenêtres avec la composante cartographique, tabulaire et graphique. La composante cartographique interactive est utilisée pour visualiser les membres de la dimension spatiale et de mesures numériques avec de simples affichages graphiques (bar, couleurs, etc.). En outre, le prototype gère aussi les mesures spatiales.

### 3.4.3 Bilan

Les solutions SOLAP existantes sont très différentes en ce qui concerne les techniques de visualisation, d'interaction et d'analyse. CommonGIS introduit, par exemple, des cartes interactives pour l'analyse des données spatio-multidimensionnelle, Polaris intègre la composante cartographique dans la table de pivot, et JMap fournit un environnement qui synchronise des représentations multiples cartographiques tabulaires et graphiques. Chaque système fournit différents opérateurs de navigation multidimensionnelle, par exemple Scotch et al. permettent de couper l'hypercube un utilisant l'opérateur de buffer et (Matias et Moura-Pires, 2007) de superposer deux hypercubes spatiaux. Enfin, les fonctionnalités d'analyse qui ne sont pas purement OLAP sont aussi très différentes. GeoMiner, par exemple, introduit des outils de fouille de données et OLAP for ArcGIS intègre aux fonctionnalités de base de l'OLAP, toutes les fonctionnalités d'analyse d'ArcGIS.

En conclusion, même si ces solutions présentent toutes des caractéristiques particulières d'analyse, de visualisation et d'interaction qui les caractérisent, elles révèlent que l'intégration des fonctionnalités OLAP et SIG dans un système unique et interactif est l'approche la plus adaptée pour une analyse spatio-multidimensionnelle effective.

## 3.5 Conclusions

L'introduction de l'information spatiale dans l'analyse multidimensionnelle implique une reformulation des concepts des entrepôts de données et de l'analyse en ligne.

L'intégration de la composante spatiale de l'information géographique en tant qu'axe ou sujet de l'analyse a mené différents auteurs à la définition respectivement de dimension spatiale et mesure spatiale.

Dans ce chapitre, nous avons détaillé ces différentes définitions et décrit les modèles conceptuels pour les bases de données spatio-multidimensionnelles proposés dans la littérature. Ainsi nous avons présenté un panorama des systèmes pour SOLAP, qui révèle que les solutions qui intègrent les fonctionnalités OLAP et SIG sont les plus adéquates à l'analyse spatiale en ligne.



---

## Partie 2 : Contributions

---





# *Chapitre 4 : De l'OLAP Spatial à l'OLAP Géographique*

Dans ce chapitre, nous redéfinissons les concepts SOLAP de dimension, hiérarchie et mesure spatiales proposés dans la littérature, pour enrichir l'analyse spatio-multidimensionnelle des caractéristiques fondamentales de l'information géographique. Ainsi, nous définissons les concepts de dimension géographique, de mesure géographique, et des nouveaux opérateurs multidimensionnels. Ces nouveaux concepts permettent d'analyser dans un contexte multidimensionnel l'information géographique dans toute sa complexité. Nous préférons le terme OLAP Géographique au terme classique de l'OLAP Spatial.

Comme nous le décrivons dans ce chapitre, l'OLAP Géographique soulève de nombreuses problématiques de modélisation, de visualisation et de navigation.

## **4.1 Motivations**

L'information géographique est la représentation d'objets ou d'un phénomène réel, localisés dans l'espace. Cette information, comme nous l'avons décrit au chapitre 2, est caractérisée par sa localisation dans l'espace, par sa forme et par ses aspects sémantiques : les propriétés descriptives alphanumériques et les relations avec d'autres objets.

La plupart des modèles SOLAP existants considèrent une vision non continue de l'espace géographique. Dans ce travail nous suivons la même approche. Nous faisons référence par le terme objet complexe, à une entité du monde réel (un patient, un produit, etc.) décrite par un ensemble d'attributs descriptifs alphanumériques (âge, nom, type, etc.) et nous utilisons le terme objet géographique pour représenter un objet complexe (une ville, un bâtiment, etc.) qui présente un attribut spatial (une géométrie). Un exemple d'objet géographique peut être la ville de Lyon qui est décrite par un nom (Lyon), par une population (1452952) et par une géométrie représentée par un point.

Les modèles SOLAP existants, bien qu'ils étendent les capacités d'analyse de l'OLAP, sous-emploient la composante sémantique de l'information géographique. De plus, ils sont caractérisés par une définition statique des données, contrairement à l'analyse spatiale. Dans cette section nous détaillons ces problématiques par rapport aux modèles des données et aux opérations de navigation.

#### 4.1.1 Modèle des données multidimensionnelles

La composante sémantique de l'information géographique influe sur la modélisation des dimensions et des mesures spatiales.

##### 4.1.1.1 Dimension spatiale

Les modèles proposés en littérature (Bédard et al., 2001) (Fidalgo et al., 2004) (Malinowski et Zimányi, 2005) (cf chap 3 § 2.1) qui introduisent l'information géographique comme axe d'analyse, sont caractérisés par la présence de l'attribut géométrique dans les membres des différents niveaux. Les hiérarchies spatiales sont définies en utilisant les attributs spatiaux ou alphanumériques des dimensions. Ces relations définissent, comme pour les hiérarchies classiques, une relation d'inclusion entre les membres de niveaux différents. Les hiérarchies spatiales sont alors des hiérarchies classiques de classification ou de spécialisation (Lujian-Mora et al., 2002) avec des attributs géométriques. Par conséquent, entre deux membres spatiaux de deux niveaux différents il existe toujours une relation topologique d'inclusion ou intersection (Malinowsky et Zimányi 2005). Ces hiérarchies représentent différentes granularités de l'information géographique, en d'autres termes, chaque niveau géographique représente une information géographique différente, par exemple les membres d'un niveau peuvent être les objets géographiques représentant les villes, et les membres d'un autre niveau peuvent représenter les départements auxquelles appartiennent les villes. Les hiérarchies spatiales sont souvent très complexes, par exemple il peut s'agir de hiérarchies non-onto, non-strictes, etc. (Malinowsky et Zimányi 2005). Pour ce type de hiérarchie, si les mesures peuvent être redistribuées sur la surface des membres, alors les relations topologiques peuvent permettre de quantifier l'apport d'un membre, par rapport à son ancêtre, dans le calcul de l'agrégation (Jensen et al., 2004) (Malinowsky et Zimányi 2005).

Un des aspects sémantiques qui caractérise l'information géographique est sa représentation à différentes échelles. Les opérateurs de généralisation, comme nous avons décrit au paragraphe 2.1.3, permettent de créer de cartes à différentes échelles et de les organiser dans une structure hiérarchique complexe. Par exemple, les régions de France peuvent être représentées à l'échelle 1:10000 et à l'échelle 1:25000. Les relations hiérarchiques de généralisation, contrairement à l'approche classique du SOLAP, ne représentent pas toujours de relations d'intersection ou inclusions entre les objets géographiques à différentes échelles. Par contre, on peut penser utiliser ce type de hiérarchie dans un modèle multidimensionnel, pour introduire dans l'analyse en ligne le concept de généralisation, qui est fondamental dans le processus d'aide à la décision spatiale. Grâce à ce type de hiérarchie, le décideur peut naviguer à travers différentes représentations à diverses échelles de la même information géographique, et « visualiser » les mesures grâce à différents degrés de précision spatiale. En revanche, pour ce type de hiérarchie, les méthodes d'agrégation ne peuvent pas être ceux des hiérarchies spatiales classiques car il est possible qu'il n'existe aucune relation topologique d'inclusion entre deux membres à différents niveaux. Il n'est pas possible alors de quantifier l'apport d'un membre par rapport à son ancêtre dans le processus d'agrégation.

De plus, dans l'analyse spatiale, les fenêtres des requêtes spatiales ne sont pas toujours conformes à des hiérarchies prédéfinies (Papadias et al., 2002). Cela implique que dans les

applications spatio-multidimensionnelles, contrairement aux approches classiques SOLAP, les hiérarchies ne sont pas figées. Ainsi, l'utilisateur doit pouvoir créer des nouveaux membres.

En somme, pour prendre en compte les aspects sémantiques et spatiaux de l'information géographique, les hiérarchies doivent intégrer les relations de généralisation et autoriser un certain degré de flexibilité et complexité, et introduire la composante spatiale en pérennant en compte les relations topologiques entre les différents membres.

#### **4.1.1.2 Mesure spatiale**

L'information géographique peut être introduite dans un modèle multidimensionnel en tant qu'axe d'analyse, mais elle peut aussi être le sujet de l'analyse. Ainsi, la composante spatiale et la composante descriptive des objets géographiques, celle-ci étant identifiée par l'ensemble des attributs alphanumériques, doivent pouvoir être utilisées comme mesures. La composante descriptive peut être utile au processus décisionnel, pour expliquer un phénomène ou caractériser un ensemble des faits. Une synthèse de l'information spatiale permet aux utilisateurs de comprendre les évolutions ou les tendances du phénomène spatial, mais seul le couplage avec la composante descriptive permet de donner une signification d'ensemble au fait géographique, de corrélérer la composante purement spatiale du phénomène à ses aspects sémantiques et d'expliquer et caractériser ce phénomène géographique par rapport aux dimensions d'analyse.

Prenons comme exemple une application multidimensionnelle concernant la supervision de la mortalité au sein des départements français. Cette application multidimensionnelle peut présenter trois dimensions : la date, l'âge de la personne et la cause du décès. La mesure est représentée par un ensemble de pointeurs aux géométries des départements. Cette application spatio-multidimensionnelle permet de répondre aux requêtes telles que : « *Où le cancer était-il la cause de plus de 10 morts parmi les personnes âgées de 50 ans chaque année ?* » Dans cette application, la composante spatiale des départements est le sujet de l'analyse multidimensionnelle, et grâce à cette mesure spatiale, l'utilisateur peut comprendre l'influence de la localisation géographique des départements sur la mortalité. Or, un département est un objet géographique décrit par plusieurs attributs descriptifs, comme la population, le nombre d'hôpitaux, le type de département (i.e. rural, industriel, etc.) et par des attributs descriptifs dérivés directement de la géométrie comme par exemple la surface. Ces attributs peuvent être utiles dans le processus décisionnel. Ils représentent une information complémentaire à la localisation du phénomène identifié par la mesure spatiale. Par exemple, le nombre d'hôpitaux et d'habitants peut refléter un système médical inadéquat. De plus, ces attributs descriptifs peuvent être utiles pour caractériser ou expliquer un phénomène, comme par exemple un lien entre une forte mortalité pour une maladie particulière et le type de département.

De plus, un objet géographique peut être en relation avec d'autres objets à travers des relations spatiales, de généralisation et non spatiales. Par exemple, un département peut présenter une relation spatiale représentant l'appartenance du département à une région, ou les départements peuvent être représentés à différentes échelles grâce aux opérateurs de généralisation qui établissent des relations (de généralisation) entre les objets géographiques à différentes échelles. Cela ouvre de nouveaux scénarii d'analyse

multidimensionnelle, car ces relations peuvent être exploitées pour changer les mesures. Par exemple, on peut reformuler l'application précédente, en exploitant la relation d'appartenance d'un département à une région pour utiliser les régions comme mesures. Alors, la requête précédente est transformée en « *Dans quelles régions (administratives) le cancer était la cause de plus de 10 morts parmi les personnes âgées de 50 ans chaque année ?* ».

Comme nous l'avons déjà mis en évidence dans le chapitre 3, de nombreux auteurs se sont intéressés à la problématique liée à l'introduction de l'information géographique en tant que mesure dans un modèle multidimensionnel. La mesure spatiale a été définie comme une collection d'objets spatiaux (Stefanovic et al., 2000), (Rivest et al., 2001), (Malinowski et Zimányi, 2004), ou comme le résultat d'un opérateur topologique ou métrique appliqué aux membres spatiaux des dimensions (Malinowski et Zimányi, 2004), (Marchand et al., 2003). Toutes ces définitions réduisent une mesure spatiale à la composante spatiale des objets géographiques et aux attributs descriptifs qui sont dérivés directement de la géométrie. La sémantique des objets géographiques, représentée par les attributs alphanumériques, n'est pas prise en compte dans le processus décisionnel.

Les deux seules approches qui cherchent à introduire cette composante descriptive dans le sujet de l'analyse sont présentées par (Fidalgo et al., 2004) et (Damiani et Spaccapietra, 2006). (Fidalgo et al., 2004) modélisent la mesure spatiale comme un membre de dimension. Cette solution n'autorise pas la modélisation des mesures comme objets géographiques car ces dernières sont modélisées comme des membres de la dimension spatiale et par conséquent elles sont agrégées sur les niveaux de la dimension selon une organisation statique définie lors de la conception de l'application multidimensionnelle, et indépendamment des autres dimensions. En d'autres termes, ces mesures sont des membres d'une dimension et donc elles ne peuvent pas être agrégées en fonction des autres dimensions. Dans le modèle présenté dans (Damiani et Spaccapietra, 2006), les membres des dimensions ne présentent aucun attribut descriptif, mais ils sont représentés par une simple valeur, par exemple un membre du niveau jour de la dimension temporelle est représenté par une simple date (i.e. 19/9/1978, 20/1/2006, etc.). Une mesure spatiale est définie comme un niveau d'une dimension. Même si une mesure spatiale œuvre à représenter un objet géographique, elle ne représente pas explicitement la composante spatiale et les attributs descriptifs et, au contraire, elle est décrite par un simple identifiant (i.e. p1, p2, etc.).

#### **4.1.2 Algèbre spatio-multidimensionnelle : flexibilité de l'analyse**

Un outil SOLAP doit intégrer les fonctionnalités OLAP et SIG dans un seul environnement. Cela signifie que, parallèlement aux opérateurs SOLAP de coupe et de forage, un outil SOLAP doit aussi fournir des fonctionnalités SIG de base et d'analyse spatiale. Dans le processus d'analyse multidimensionnelle, les chemins d'analyse sont imprédictibles, mais le contexte d'analyse, représenté par les données et le modèle multidimensionnel (dimensions et mesures), est défini lors de la conception de l'entrepôt de données. L'utilisateur navigue dans l'hypercube à travers les hiérarchies des dimensions, en comparant les mesures, qui sont agrégées à différentes granularités avec des fonctions d'agrégation. Contrairement à l'approche utilisée dans le processus d'analyse multidimensionnelle, l'analyse spatiale est flexible et itérative. Le processus d'analyse

spatiale, comme nous l'avons décrit au chapitre 2, est un processus itératif où les données et les méthodes d'analyse utilisées sont modifiées à chaque itération. Les données géographiques, grâce aux méthodes d'analyse spatiale de transformation (i.e., le buffer, l'overlay, etc.), sont modifiées ou remplacées tout au long du processus d'analyse. Pour ce faire, nous considérons comme nécessaire l'introduction et l'adaptation des opérateurs d'analyse spatiale dans un contexte OLAP, afin d'intégrer la flexibilité de l'analyse spatiale au modèle multidimensionnel. L'adaptation de ces opérateurs à l'OLAP implique que les données du modèle multidimensionnel doivent être flexibles, ou en d'autres termes, que les dimensions qui incluent l'information géographique doivent pouvoir être reformulées selon l'exigence de l'utilisateur et, ce qui pose plus de problèmes, que les mesures doivent alors être recalculées.

Les modèles SOLAP existant utilisent des modèles de données différents pour la représentation des membres et des mesures spatiales. En terme d'analyse multidimensionnelle, cela se traduit par une asymétrie entre la mesure et la dimension spatiale, ce qui oblige l'utilisateur à concevoir deux applications différentes qui permettent de répondre à des requêtes multidimensionnelles diverses : l'une avec la mesure spatiale et l'autre avec une dimension spatiale. Ce manque de symétrie et de flexibilité représente une limite importante des solutions SOLAP existantes, car l'information géographique peut être aussi bien incluse en dimension qu'en mesure. En effet, dans les applications multidimensionnelles, une vision symétrique des membres des dimensions et des mesures permet de changer à la volée le modèle de l'application en augmentant les capacités d'analyse du SOLAP classique. Prenons comme exemple une application concernant la pollution de l'air des départements de France. Une application multidimensionnelle possible présente une dimension temporelle, une dimension représentant les polluants (i.e. le zinc, le fer, etc.) et une dimension représentant la valeur de la pollution (i.e. 0-4 mg/l, 5-9 mg/l, etc.) et comme mesure les objets géographiques représentant les départements de France. Cette application permet de répondre aux questions telles que : « *Quels sont les départements pollués au zinc avec une valeur de pollution entre 0 et 4 mg/l chaque jour ?* ». Si l'on considère que les mesures et les membres des dimensions sont représentés à travers les mêmes modèles de données, alors, à partir des mêmes données de la table des faits, on pourra avoir une application où la valeur de la pollution est la mesure. Elle permettra de répondre à ce type de requêtes « *Quelle est la valeur maximale de pollution du zinc pour chaque département et pour chaque mois ?* ».

## 4.2 Contexte

Dans cette section nous allons décrire la notation graphique que l'on utilisera pour représenter visuellement les modèles spatio-multidimensionnelles et le contexte de notre travail.

### 4.2.1 Cas d'étude : la pollution des eaux de la lagune de Venise

Notre travail de thèse s'inscrit dans le cadre d'une collaboration internationale avec l'organisation italienne CORILA (Consorzio per la Gestione del Centro di Coordinamento delle Attività di Ricerca inerenti il Sistema Lagunare di Venezia). CORILA a pour partenaire l'université Cà Foscari de Venise, l'Institut Universitaire d'Architecture de Venise, l'Université de Padoue et le Conseil National de la Recherche italienne. Elle est

sous la tutelle du Ministère de l'Université et la Recherche Scientifique et Technologique d'Italie. Le but de cette organisation est la sauvegarde environnementale, architecturale et économique de la lagune de Venise. CORILA joue un rôle clé dans les activités de recherche concernant ces problématiques et fait aussi fonction de médiateur avec d'autres organisations internationales de recherche qui interviennent dans les grandes lignes thématiques de recherche.

Notre équipe intervient dans l'axe de recherche « gestion des données ». L'objectif est de fournir aux spécialistes environnementaux (biologistes, géologues, etc.) un système pour l'analyse spatio-multidimensionnelle de données environnementales.

Ces données concernent la pollution des eaux de la lagune et sont issues d'une campagne de suivi appelée « Drain » (Bimonte et al., 2006d). Pendant cette campagne, les experts ont produit huit fichiers qui contiennent les mesures de la pollution en 25 différentes zones de la lagune (Figure 4.1), à 2826 moments différents (3 années, 28 mois et 207 jours), et pour 100 polluants. Notre mission est de fournir un outil qui permet d'analyser ces données en utilisant le paradigme multidimensionnel. Plus particulièrement, il s'agit de réaliser un système pour l'analyse spatio-multidimensionnelle des zones de lagune polluées en fonction du type de polluant, du temps et de la valeur de pollution, aussi bien que des valeurs de la pollution en fonction du temps, de polluants et des zones de la lagune. Dans le reste de ce travail de thèse, nous utilisons cette application concernant la pollution de la lagune de Venise pour décrire nos solutions théoriques et d'implémentation.

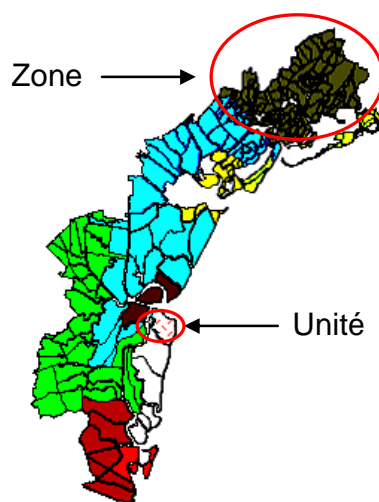


Figure 4.1. Carte de la lagune de Venise.

#### 4.2.2 Notation graphique pour les modèles conceptuels applicatifs

Dans le reste du chapitre, nous utilisons la notation graphique du modèle conceptuel pour les entrepôts de données spatiales Spatial MultiDim (Malinowsky, 2006), décrit au chapitre 3, auquel nous avons apporté quelques modifications. La Figure 4.2 expose notre extension pour représenter une mesure géographique. Nous notons « / » un attribut dérivé.

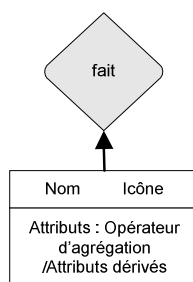


Figure 4.2. Extension de la notation graphique pour les entrepôts de données spatiales présentée dans (Malinowsky, 2006) avec une mesure géographique et des attributs dérivés

### 4.3 Concepts principaux de l'OLAP Géographique

L'analyse faite dans la section précédente révèle que les modèles SOLAP existants ne prennent peu ou pas en compte l'aspect sémantique de l'information géographique dans l'analyse multidimensionnelle. Nous proposons de reformuler les principaux concepts du SOLAP, pour définir le nouveau concept d'OLAP Géographique.

L'OLAP Géographique vise à introduire la composante spatiale et la composante sémantique de l'information géographique dans l'analyse multidimensionnelle. En particulier, en se focalisant sur la composante descriptive des objets géographiques, l'OLAP Géographique reformule le concept de mesure spatiale. La mesure n'est alors pas simplement une collection d'objets spatiaux, mais un objet géographique qui peut être associé à d'autres objets géographiques ou non. L'OLAP Géographique reformule aussi le concept de dimension spatiale. Une dimension permet de modéliser les relations spatiales, descriptives et de généralisation entre les objets géographiques et/ou complexes. Elle est caractérisée par une flexibilité et une complexité en termes des données. Enfin, ces nouveaux concepts permettent d'élargir les opérateurs classiques du SOLAP et autorisent de nouvelles formes d'analyse spatio-multidimensionnelle. Les opérateurs d'analyse spatiale et une vision symétrique des mesures et des dimensions sont aussi pris en compte dans l'OLAP Géographique.

#### 4.3.1 Dimension géographique

##### *Définition 1. Dimension géographique*

*Une dimension est dite «géographique» si les membres d'au moins un niveau sont des objets géographiques.*

Dans une dimension géographique, il est possible que tous les membres présentent l'attribut spatial ou que seuls quelques niveaux incluent l'attribut spatial.

Une dimension géographique structure l'information à différentes granularités représentées par ses niveaux, dont les membres peuvent être liés par des relations spatiales, des relations de généralisation et des relations descriptives.

### 4.3.1.1 Hiérarchie descriptive

Une hiérarchie descriptive organise l'information géographique à différentes granularités sémantiques.

#### Définition 2. Hiérarchie descriptive

Une « hiérarchie descriptive » d'une dimension géographique est une hiérarchie OLAP de classification ou de spécialisation : elle est définie en utilisant les attributs descriptifs des objets.

Un exemple de hiérarchie descriptive pour la lagune de Venise est montré en Figure 4.3. Cette hiérarchie classe les unités de la lagune par rapport au type d'unité (marais, portuaire, etc.). Le niveau « Type » de cette hiérarchie descriptive identifie des regroupements d'unités de la lagune qui sont caractérisés par le même type.

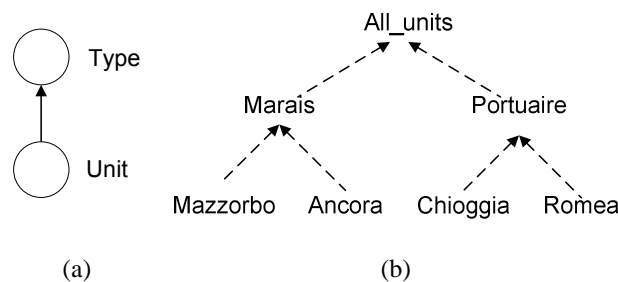


Figure 4.3. Hiérarchie descriptive a) Schéma b) Instance.

L'application multidimensionnelle qui utilise cette hiérarchie descriptive est montrée en Figure 4.4.

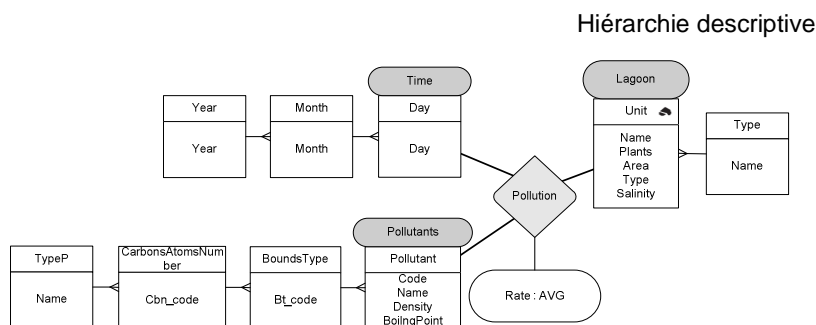


Figure 4.4. Application multidimensionnelle avec une dimension géographique décrite par une hiérarchie descriptive.

Les dimensions sont le temps, les polluants et une dimension géographique qui contient la hiérarchie descriptive. Nous notons que les niveaux d'une dimension géographique peuvent être géographiques mais aussi alphanumériques, comme le niveau « Type » de la dimension « Lagoon ». Cette application permet d'analyser la pollution des eaux de la



lagune en fonction du temps, des polluants et des zones de la lagune. Une requête multidimensionnelle peut être : « *Quelle est la valeur moyenne de la pollution pour les zones de type marais pour chaque année et pour les polluants de type organique ?* ».

### 4.3.1.2 Hiérarchie spatiale

Une hiérarchie spatiale organise l'espace géographique dans une structure hiérarchique où chaque niveau représente une information géographique différente.

#### Définition 3. Hiérarchie spatiale

*Une « hiérarchie spatiale » d'une dimension géographique est une hiérarchie où les membres de différents niveaux sont liés par des relations topologiques d'inclusion et/ou intersection.*

Cette hiérarchie peut éventuellement être automatiquement calculée grâce à l'attribut géométrique. De plus, si les mesures peuvent être redistribuées sur la surface des membres, les relations topologiques qui caractérisent cette hiérarchie peuvent permettre de quantifier l'apport d'un membre, par rapport à son ancêtre, dans le calcul de l'agrégation.

Un exemple de schéma et d'instance d'une hiérarchie spatiale est montré en Figure 4.5 où plusieurs unités de la lagune de Venise « Unit » sont regroupées en plusieurs régions « Zone ». Les représentations cartographiques des deux niveaux géographiques « Unit » et « Zone » sont représentées respectivement en Figure 4.1. Une relation topologique d'inclusion existe entre les « Units » et les « Zones ». La géométrie d'une zone est l'union topologique des géométries d'un ensemble d'unités.

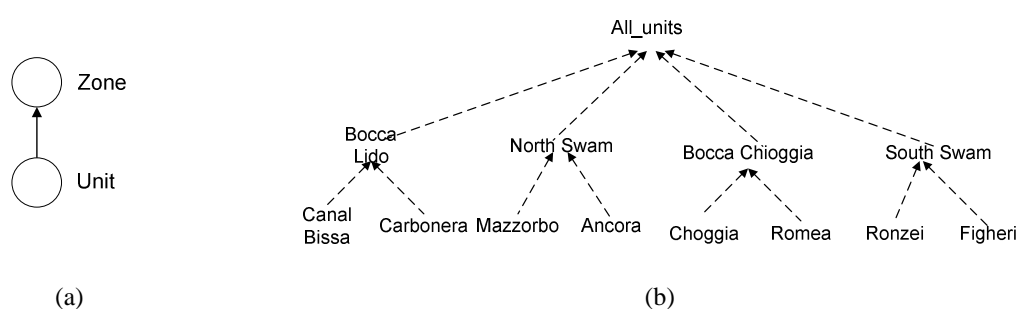


Figure 4.5. Hiérarchie spatiale a) Schéma b) Instance.

Un exemple d'application multidimensionnelle qui utilise cette hiérarchie est montré en Figure 4.6.

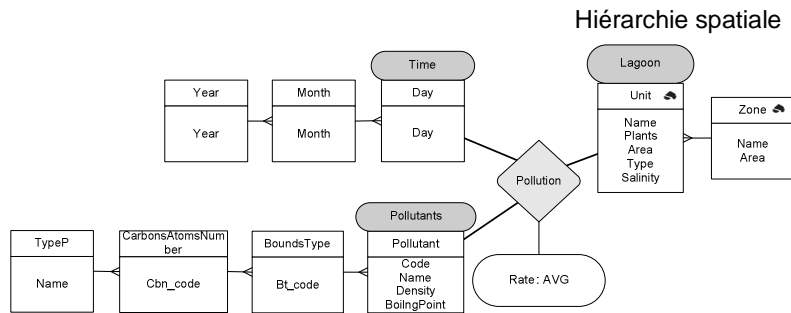


Figure 4.6. Application multidimensionnelle avec une dimension géographique décrite par une hiérarchie spatiale.

Cette application permet de répondre aux requêtes telles que : « *Quelle est la valeur moyenne de la pollution pour chaque zone de la lagune en septembre 2000 et pour chaque polluant ?* ».

#### 4.3.1.3 Hiérarchie de généralisation

Une hiérarchie de généralisation représente un ensemble de couches qui décrivent la même information géographique à différentes échelles.

##### **Définition 4. Hiérarchie de généralisation**

*Une « hiérarchie de généralisation » d'une dimension géographique est une hiérarchie où les membres des niveaux représentent la même information géographique à différentes échelles et dont les membres d'un niveau sont les résultats de la généralisation des membres du niveau directement inférieur.*

Dans la même application multidimensionnelle que celle de la Figure 4.5, nous substituons à la hiérarchie spatiale une hiérarchie de généralisation qui représente les unités de lagune de Venise à deux différentes échelles : 1 : 1000 et 1 : 500 (Figure 4.7).

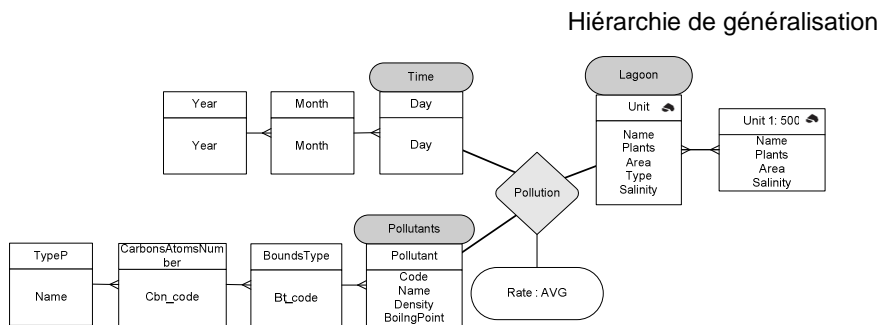


Figure 4.7. Application multidimensionnelle avec une dimension géographique décrite par une hiérarchie de généralisation

La Figure 4.8 montre un exemple de cette hiérarchie et les deux cartes associées. La carte généralisée est obtenue en utilisant les opérateurs de généralisation d'*agrégation*, de *simplification* et de *sélection* (cf chap 2 § 1.3). La géométrie de « Saca Ghebo Storto » est simplifiée, « Treporti » et « Botta Sora Canal » sont fusionnées dans une seule grande zone dont la géométrie est simplifiée et « Paleazza » est éliminée.

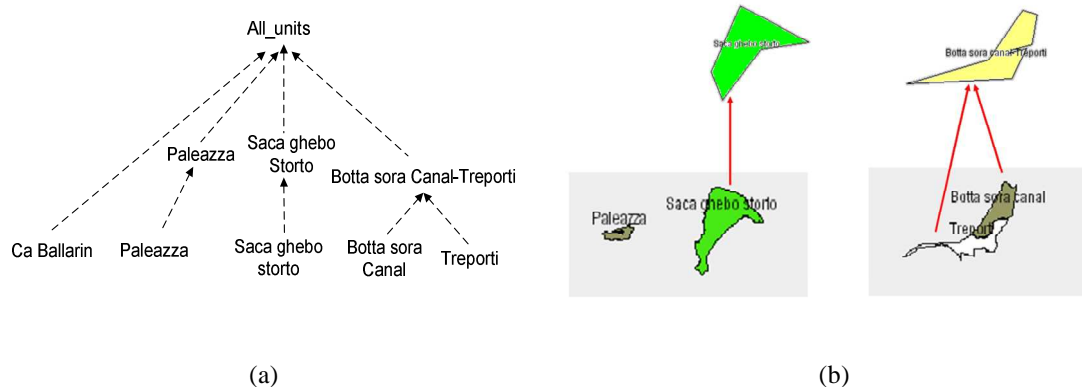


Figure 4.8. Hiérarchie de généralisation a) Instance b) Représentation cartographique.

Cette application multidimensionnelle permet de répondre aux questions telles que : « *Quelle est la pollution moyenne pour polluant, pour chaque année et pour chaque zone de la lagune à l'échelle 1 : 500 ?* »

En utilisant ce type de hiérarchie, l'application spatio-multidimensionnelle s'enrichit du pouvoir d'analyse associé à la généralisation cartographique. Une représentation des mesures à travers des cartes à différentes échelles permet à l'utilisateur d'avoir un aperçu visuel global et simplifié du phénomène, en excluant les informations n'étant pas primordiales pour la compréhension de ses caractéristiques générales. Par voie de conséquence, une hiérarchie de généralisation autorise l'utilisateur à employer au mieux ses capacités de perception visuelle.

#### 4.3.1.4 Problématiques

Différentes problématiques sont soulevées par la modélisation et la visualisation des dimensions géographiques.

##### 4.3.1.4.1 Modélisation

Dans une hiérarchie spatiale, le calcul de l'agrégation doit prendre en compte les relations topologiques d'inclusion ou d'intersection entre les membres. En effet, puisque les mesures peuvent dépendre des géométries des membres, ces relations peuvent affecter l'exactitude du calcul d'agrégation. La représentation de ce degré d'inclusion topologique, qui, à la différence des modèles classiques peut être partiel (Jensen, et al., 2004), est nécessaire dans une application multidimensionnelle avec une hiérarchie spatiale.

L'information géographique peut être organisée à différentes échelles grâce à des hiérarchies complexes où les membres à différents niveaux ne sont pas liés forcément par

de simples relations topologiques d'inclusion ou d'intersection. Par conséquent, est-il possible d'utiliser les fonctions d'agrégation classiques (i.e. SUM, MIN, MAX) sans prendre en compte les aspects topologiques, comme dans le cas des hiérarchies spatiales ? Quelles méthodes d'agrégation doit-on utiliser pour représenter l'information représentée par les mesures lors du changement d'échelle ? La définition des agrégations possibles sur ces hiérarchies reste un défi majeur. De plus, ces hiérarchies sont très complexes et les membres de différents niveaux peuvent être liés par une relation de multi-association (Spaccapietra, et al., 2007), qui établit un lien entre deux groupes d'objets géographiques. Ces relations impliquent que pour ce type d'applications multidimensionnelles, un degré d'imprécision doit être permis lors d'une opération d'agrégation, car l'apport d'un membre lors de l'agrégation sur son père ne peut pas être quantifié. Pour expliquer cette problématique nous présentons un exemple concernant les accidents domestiques. Les dimensions sont la dimension temporelle, une dimension représentant le type d'accident (i.e. électricité, gaz, etc.) et une dimension géographique avec une hiérarchie de généralisation représentant les bâtiments où ces accidents ont eu lieu à deux échelles différentes (Figure 4.9a). Le fait est décrit par une mesure numérique qui est le nombre d'accidents, qui est agrégée banalement en utilisant la somme. Dans la hiérarchie de généralisation, cinq bâtiments « bat1 », « bat2 », « bat3 », « bat4 », « bat5 » sont liés à 3 bâtiments à l'échelle moins détaillée « batA », « batB » et « batC » comme montré en Figure 4.9b. Supposons par exemple que pour chaque bâtiment « bat1 », « bat2 », « bat3 », « bat4 », « bat5 » il y eu 10 accidents. Puisque les bâtiments « batA », « batB » et « batC » sont obtenus en utilisant une opération de généralisation de *typification* (cf chap 2 § 1.3), alors on ne peut pas dire exactement quels bâtiments à l'échelle plus détaillée ont été fusionnés. Par exemple, on ne peut pas affirmer que les bâtiments « bat1 » et « bat2 » forment le « batA », car l'opération de typification donne lieu à une relation de multi-association. Lors d'une requête multidimensionnelle portant sur le niveau moins détaillé de la hiérarchie de généralisation, on ne peut pas appliquer la somme pour calculer le nombre d'accidents pour le bâtiment « batA ». En revanche, on peut dire que 50 accidents (la somme des accidents pour les bâtiments « bat1 », « bat2 », « bat3 », « bat4 », « bat5 ») ont eu lieu dans les bâtiments « batA », « batB » et « batC ». Les relations topologiques entre les bâtiments à différentes échelles ne peuvent pas être utilisées dans le processus d'agrégation, car les bâtiments ne sont pas liés par des relations topologiques d'inclusion et intersection.

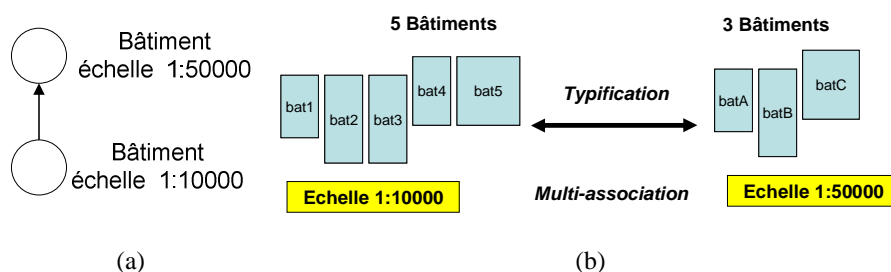


Figure 4.9. Hiérarchie de généralisation a) Schéma b) Instances et relation multi-association.

#### 4.3.1.4.2 Visualisation

La visualisation est une des caractéristiques la plus importante et en même temps une des plus problématiques d'un outil SOLAP (Rivest et al., 2005). L'interface utilisateur doit supporter plusieurs dimensions géographiques, des cartes qui mixent différents types spatiaux (point, polygone, ligne) et doit permettre de naviguer entre les niveaux textuels et cartographiques. De plus, elle doit synchroniser des composantes graphiques, tabulaires et cartographiques. La synchronisation est une tâche compliquée car ces composantes ont toutes des pouvoirs d'expression différents. Par exemple, dans la hiérarchie spatiale représentant la subdivision administrative de France, puisque la géométrie d'une région recouvre les géométries de ses départements, ces objets géographiques ne peuvent pas être affichés en même temps sur une carte, contrairement à la table de pivot. De plus, de nombreux problèmes de sémiologie doivent être résolus pour aboutir à une visualisation correcte des mesures sur la dimension géographique (Soumahoro, 2005). La façon dont ces mesures doivent être visualisées sur la carte dépend du nombre et de la nature des mesures (quantitative ou qualitative), et de la représentation courante (i.e. point, ligne, polygone) des membres de la dimension géographique affichés sur la carte.

L'interaction avec la composante cartographique représente aussi un défi majeur. Les opérateurs de navigation doivent être adaptés au type de hiérarchie. Par exemple, dans les hiérarchies de généralisation, contrairement aux hiérarchies spatiales et descriptives, une opération de forage ne peut pas être appliquée à un seul membre, car la carte résultante de cette opération mélangerait des objets géographiques à différentes échelles.

#### 4.3.2 Mesure géographique

Une mesure géographique introduit l'information géographique en tant que sujet d'analyse dans une application multidimensionnelle.

##### 4.3.2.1 Définition et application

Nous définissons une mesure géographique de la façon suivante :

##### ***Définition 5. Mesure géographique***

*Une mesure est dite géographique si elle est un objet géographique qui, comme un niveau d'une dimension géographique, peut appartenir à un ou plusieurs schémas hiérarchiques.*

Un exemple d'application qui utilise une mesure géographique est montré en Figure 4.10. Les dimensions sont le temps (« Time »), les polluants (« Pollutants ») et des valeurs de pollution regroupées par intervalles (« Rate »). La mesure est l'objet géographique « Unit » qui représente une unité de la lagune. Elle est décrite par attributs descriptifs : un nom (Murano, Chioggia, etc.), une liste de plantes (Spartima Marittima, etc.), un type (industrielle, agricole, etc.), un index de salinité (24, 27, etc.) et la surface, et un attribut spatial. Ce modèle permet d'analyser les unités de la lagune en fonction du temps, de polluants et des valeurs de pollution, et de répondre aux questions telles que : « *Quelles sont les zones de lagune, leur salinité, leur plantes qui étaient fortement polluées au zinc,*

*pour chaque année ? » ou « Quelles sont les unités de lagune, polluées au zinc, pour chaque jour et pour chaque valeur de pollution (Rate5) ? ».*

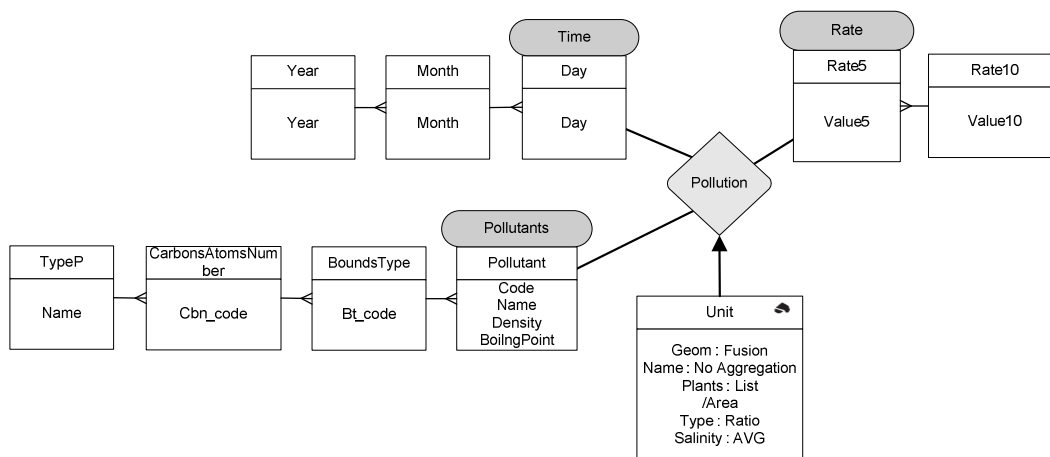


Figure 4.10. Mesure géographique « Unit ».

L'agrégation d'une unité de la lagune correspond à l'agrégation de ses attributs descriptifs et de sa composante spatiale. Plus précisément, dans cette application, nous choisissons de n'appliquer aucune agrégation aux noms, de fusionner les géométries, de créer une liste pour les plantes, de moyenner la salinité et pour le type, d'utiliser une moyenne sur la surface. Nous notons que la surface n'est pas agrégée car elle peut être calculée en utilisant la géométrie résultante de l'agrégation spatiale. Ces attributs descriptifs peuvent être utiles au processus décisionnel, par exemple un index de salinité très élevé peut se révéler toujours associé à une forte pollution au fer.

Une mesure géographique est un objet géographique du monde réel comme un membre d'une dimension géographique. Une mesure peut donc appartenir à une structure hiérarchique. Par exemple, une unité peut appartenir à la hiérarchie spatiale avec les niveaux « Unit » et « Zone » ou encore une unité peut être représentée à deux différentes échelles et appartenir à une hiérarchie de généralisation. Par conséquent, l'information géographique qui représente le sujet de l'analyse peut être analysée à différentes granularités qui correspondent aux niveaux des hiérarchies. A chaque granularité de la mesure correspond une requête multidimensionnelle différente. Si l'on utilise la granularité « Zone », par exemple, une requête possible est « *Quelles sont les zones polluées par le zinc pour chaque jour, avec une valeur de 0-4 mg/l ?* » ou encore, afin d'avoir un aperçu plus ample du phénomène géographique on peut utiliser les unités à l'échelle 1 : 500 et répondre aux requêtes telles que : « *Quelles sont les unités à l'échelle 1 : 500 polluées au zinc pour chaque jour, avec une valeur de 0-4 mg/l ?* » (Figure 4.11).

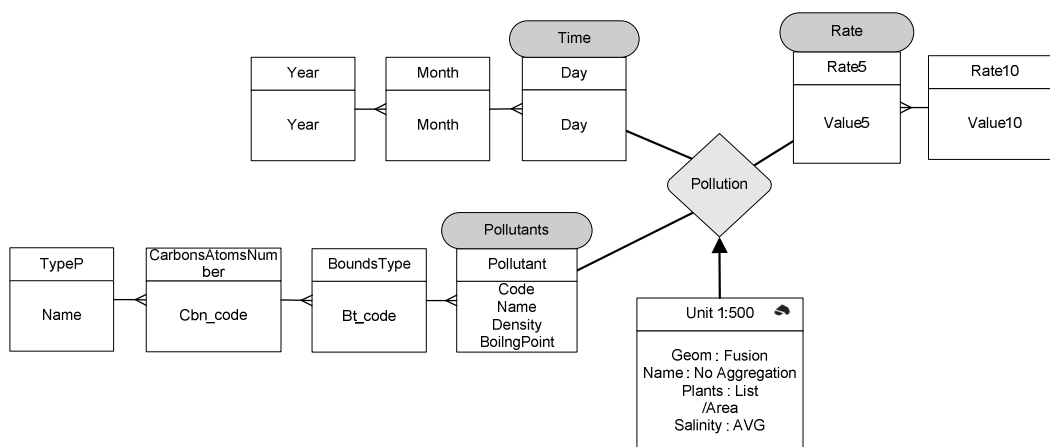


Figure 4.11. Mesure géographique « Unit » à l'échelle 1 : 500.

Cette vision de la mesure comme un objet géographique en relation avec d'autres objets géographiques nous permet d'exploiter l'aspect spatial aussi bien que la composante sémantique (représentée par les attributs descriptifs et les relations avec les autres objets) de l'information géographique lorsqu'elle représente le sujet de l'analyse spatio-multidimensionnelle.

Dans les modèles OLAP classiques, un fait est un concept relevant du processus décisionnel qui représente le phénomène analysé et peut-être décrit par un ensemble de valeurs numériques appelées mesures (Kimball, 1996). Dans le reste de ce travail de thèse, nous utiliserons le terme de « mesure » pour identifier une mesure complexe composée par plusieurs attributs. Par conséquent, en faisant un parallèle avec l'approche classique, notre concept de mesure correspondra à un fait et un attribut à une mesure.

#### 4.3.2.2 Problématiques

Les mesures géographiques soulèvent différentes problématiques liées à la modélisation et à l'implémentation.

##### 4.3.2.2.1 Modélisation

Dans les modèles OLAP classiques, une mesure représente une description d'un fait. Sa sémantique est réduite à une simple description quantitative (i.e. un montant, une quantité, etc.). Une mesure ou un ensemble de mesures n'identifient pas une entité particulière du monde réel. En revanche, dans notre cas d'étude, la mesure est l'objet géographique « Unit » qui est une entité du monde réel. Une « Unit » est identifiée par l'ensemble de ces attributs alphanumérique et une géométrie. Cette mesure est fortement typée et elle est en relation avec d'autres objets, à travers des relations spatiales, de généralisation et non-spatiales. Ce type de mesure ne peut pas être représenté, comme dans le cas classique, à travers les attributs de la relation qui représente le fait. Il s'agit d'un fait modélisé par une relation de type plusieurs à plusieurs entre les dimensions et l'entité géographique représentant la mesure.

Cette nouvelle conception de la mesure comme un objet géographique soulève de nombreuses problématiques.

Le concept de mesure tel que nous la proposons et son rôle dans la structuration multidimensionnelle sont complètement différents de l'approche classique. La façon de représenter les attributs alphanumériques et géométriques, et de les lier afin qu'ils aient une signification comme un tout, et la façon d'inclure une mesure dans une hiérarchie représentent des défis majeurs.

L'agrégation de ces mesures est un important problème. Le processus d'agrégation peut utiliser plusieurs attributs en même temps, contrairement aux agrégations SQL classiques. De plus, le modèle doit agréger les attributs d'un objet géographiques. La manière dont ces attributs doivent être agrégés et comment les fonctions d'agrégation doivent prendre en compte la dépendance sémantique qui lie ces attributs soulèvent des questions. Nous illustrons cette dernière problématique avec un simple exemple d'application concernant la supervision des zones d'un parc touchées par des phénomènes naturels comme les feux, les avalanches, etc. L'application spatio-multidimensionnelle présente deux dimensions : le temps et le type de phénomène. La mesure est représentée par la zone du parc touchée. Chaque zone est décrite par une géométrie, la surface et le nombre de plantes endommagées. L'agrégation des mesures géographiques représentant les zones du parc correspond à l'agrégation de leurs géométries et des plantes endommagées. Pour la composante géométrique, l'utilisation de deux fonctions d'agrégations spatiales différentes est possible : l'union et l'intersection. L'utilisation de l'union nous permet de répondre à des requêtes multidimensionnelles telles que : « *Où et combien de plantes ont été endommagées par au moins un feu en 2000 ?* ». Si l'on utilise l'intersection alors la même requête devient : « *Où et combien de plantes ont été endommagées par tous les feux en 2000 ?* ». Une fois la fonction d'agrégation spatiale établie, le problème est de savoir comment agréger ou recalculer le nombre de plantes pour la nouvelle géométrie. Dans la première application qui utilise l'union, la somme a du sens. En revanche, l'utilisation de l'intersection implique que la somme ne peut pas être appliquée car elle n'est pas représentative du nombre de plantes dans la zone générée par l'intersection. En d'autres termes, cela signifie que le concept d'additivité de l'OLAP classique doit être repensé lorsque la mesure est un objet géographique. Les fonctions d'agrégation dépendent du type de d'attribut de la mesure géographique et des dimensions utilisées comme dans les modèles des bases de données multidimensionnelles classiques, mais aussi de l'agrégation spatiale appliquée.

Enfin, puisque la mesure est incluse dans une structure hiérarchique, comment peut-on exploiter la vision de la mesure à différentes granularités dans le processus décisionnel ? Est-il suffisant d'autoriser une navigation entre différents hypercubes définis a priori, un pour chaque granularité différente de la mesure, comme dans le modèle présenté en (Damiani et Spaccapietra, 2006) ? Faut-il introduire des nouveaux opérateurs multidimensionnels qui changent la granularité de la mesure en utilisant les données résultant des opérateurs de navigation multidimensionnelle ?



#### 4.3.2.2 Visualisation

Les clients SOLAP sont basés sur l'intégration des tables de pivot, des affichages graphiques et des cartes interactives. Ces composantes permettent de naviguer dans l'hypercube spatial et de visualiser, comparer et comprendre les données. Les cartes affichent les attributs numériques des mesures à travers différents affichages.

Quand les mesures ne sont pas numériques, mais qu'elles représentent des données géographiques, une simple description textuelle n'est pas suffisante pour le processus d'analyse car le pouvoir expressif des cartes n'est pas exploité. Dans un hypercube avec une mesure géographique, chaque cellule peut être vue comme une carte. Une solution SOLAP qui rend « visibles » les mesures géographiques, doit fusionner dans un seul environnement toutes les méthodes OLAP de visualisation et d'interaction et des techniques de géovisualisation. Cela, dans le but de garantir le même pouvoir expressif de la table de pivot et de ses affichages graphiques dans le cas où le sujet d'analyse est l'information géographique, pour permettre de comparer des données géographiques et donc de comprendre leurs relations spatiales et alphanumériques. Pour présenter les caractéristiques fondamentales de l'interface du client, présentons tout d'abord une application spatio-multidimensionnelle concernant l'étude de la mortalité. Elle présente trois dimensions : le temps, les maladies et une dimension géographique. La dimension géographique est décrite par la hiérarchie région et pays. La mesure est l'objet géographique « département ». Le modèle permet d'analyser les départements touchés par des maladies en fonctions du temps, des maladies, de la région et du pays d'appartenance.

L'interface doit alors :

- Visualiser les membres et les mesures géographiques en même temps. La mesure « département » est analysée en fonction des régions.
- Visualiser les relations spatiales entre mesures et dimensions. Les départements touchés par une maladie peuvent être situés dans une zone particulière d'un pays ou d'une région.
- Utiliser des techniques de géovisualisation pour visualiser les relations spatiales entre les différentes mesures géographiques. Comprendre l'évolution spatiale du SIDA dans chaque région au cours de 10 dernières années.
- Afficher les attributs thématiques des objets géographiques. Rendre visible l'information relative à la population pour identifier les zones les plus à risque.
- Représenter visuellement le modèle conceptuel de l'application multidimensionnelle.
- Rendre accessibles les opérateurs OLAP à travers l'interaction avec l'interface visuelle.

(Gatalsky et al., 2004) montrent que pour l'analyse de l'information spatio-temporelle, des techniques avancées de géovisualisation sont nécessaires. De la même façon, les

métaphores classiques de visualisation cartographique comme par exemple l'overlay ou le MultiMaps, ne sont pas adaptées au paradigme spatio-multidimensionnel. En effet, la représentation à travers l'overlay du fait que la même mesure géographique corresponde à plusieurs combinaisons des membres de dimensions (par exemple que le département du Rhône est touché par le SIDA en 1978 et 2000) est une opération très complexe. Une solution possible est d'utiliser des variables visuelles (forme, couleur et orientation) adaptées aux valeurs qualitatives (Bertin et Bonin, 1992), pour représenter les coordonnées des cellules qui incluent la mesure géographique. Par exemple, en utilisant la couleur, un département est rouge s'il a été touché par le SIDA en 1978 et 2000, vert en 1980 et 1981, etc. Ces variables visuelles ne sont pas adaptées pour rendre « visible » une information aussi complexe comme l'information spatio-multidimensionnelle car le nombre de cellules à visualiser est très élevé. De plus, l'utilisation de variables visuelles pour représenter les coordonnées des cellules associées à la carte diminue le nombre de possibilités de visualisation des attributs descriptifs. L'overlay, même s'il révèle les relations spatiales entre différentes mesures géographiques, n'est pas adapté pour la visualisation de données spatio-multidimensionnelles. Le MultiMaps, qui est une grille des cartes thématiques, quant à lui est conçu pour mettre en évidence les différences entre les attributs alphanumériques plutôt que les relations spatiales.

En conclusion, de nouvelles méthodes de visualisation et d'interaction, intégrant les techniques de visualisation de l'OLAP et de géovisualisation, sont nécessaires pour les applications multidimensionnelles dont le sujet d'analyse est l'information géographique.

### 4.3.3 Opérateurs d'analyse spatio-multidimensionnelle

Au chapitre 3, nous avons illustré les opérateurs d'analyse spatio-multidimensionnelle de forage et de coupe. Ils permettent respectivement de monter ou descendre dans les hiérarchies de dimensions, et de couper l'hypercube en utilisant l'information spatiale.

Par la seule action de ces opérateurs, le paradigme d'analyse spatio-multidimensionnel classique n'est pas suffisamment flexible en termes de données utilisées dans le processus décisionnel, pour exploiter les aspects sémantiques de l'information géographique, et pour permettre d'introduire dans un contexte multidimensionnel les méthodes d'analyse spatiale. Or, la possibilité d'introduire dynamiquement de nouveaux membres de dimensions, d'intervertir sujet et axe d'analyse et de changer la granularité de la mesure, augmente et complète les capacités d'analyse des modèles SOLAP classiques. Par conséquent, en plus des opérateurs de forage et coupe, nous identifions trois nouveaux types d'opérateurs d'analyse spatio-multidimensionnelle :

- Les opérateurs qui modifient dynamiquement la structure de l'hypercube.
- L'opérateur qui permet d'intervertir mesures et dimensions.
- Les opérateurs de navigation dans la hiérarchie de la mesure.

Nous allons décrire ces nouveaux opérateurs et les problématiques associées dans le reste de cette section.

#### 4.3.3.1 Modification dynamique de la structure de l'hypercube

Parallèlement aux opérateurs multidimensionnels classiques de forage et de coupe, nous définissons les opérateurs qui modifient la structure de l'hypercube comme des opérateurs qui permettent à l'utilisateur de créer de nouveaux membres à la volée grâce aux opérateurs d'analyse spatiale.

Un exemple de ce type d'opérateur peut être représenté par l'adaptation de l'overlay au paradigme multidimensionnel. L'overlay est un opérateur d'analyse spatiale de transformation qui permet de mettre en relation des informations issues de sources variées (recensement, relevés de terrain etc.) et de nature différente (réseau routier, bâtiments, etc.). Il prend en entrée deux cartes et il génère une carte dont les géométries des objets géographiques sont recalculées grâce à l'opération topologique d'intersection. L'ensemble des objets géographiques de la carte résultante dépend aussi de l'opérateur logique utilisé (AND ou OR). Un exemple d'overlay qui utilise l'opérateur logique AND est montré en Figure 4.12.

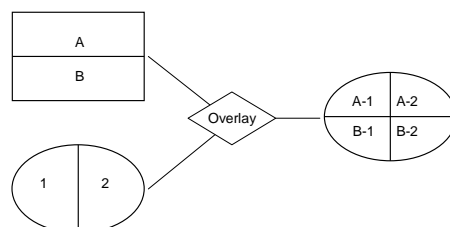


Figure 4.12. Overlay utilisant l'opérateur logique AND.

Dans un contexte spatio-multidimensionnel, il est envisageable d'appliquer l'overlay entre la carte qui représente un niveau d'une dimension géographique et une autre couche choisie par l'utilisateur. Le résultat de cette opération crée de nouveaux membres de la dimension géographique. Pour ces nouveaux membres, les attributs des mesures doivent être recalculés en utilisant les parties de membres issues de l'opération d'overlay, les membres originaux des deux couches et les valeurs des attributs des mesures dans la table de faits associés.

Contrairement aux opérateurs de forage et de coupe, ces opérateurs représentent une approche toute nouvelle dans l'analyse spatio-multidimensionnelle. En effet, les opérateurs d'analyse spatiale SIG utilisés dans les solutions SOLAP existantes, comme l'opérateur de buffer présenté dans (Scotch et Parmanto, 2006), ne sont pas traduits dans une vision multidimensionnelle. Ils permettent uniquement de naviguer dans l'hypercube à partir de la composante spatiale où ils représentent des fonctionnalités « orientée » SIG orthogonales aux fonctionnalités SOLAP. A travers les opérateurs qui modifient la structure de l'hypercube, le paradigme multidimensionnel gagne en flexibilité et en capacités d'analyse.

#### 4.3.3.2 Permutation entre mesure et dimension

Dans notre approche, l'information géographique peut être présente en mesure comme en dimension. Elle est modélisée dans les deux cas comme un ensemble d'objets géographiques. L'utilisation d'une mesure géographique ou d'une dimension géographique

donne lieu à des modèles applicatifs différents comme nous l'avons montré dans les sections précédentes. D'où, l'intérêt de pouvoir intervertir dimension et mesure, dynamiquement, pendant les processus d'analyse multidimensionnelle.

Transformer une mesure géographique en dimension, implique qu'un objet géographique ou complexe devient la nouvelle mesure. Dans notre approche, une mesure est définie comme un objet géographique et par conséquent, comme un objet complexe représenté par un ensemble d'attributs ayant une signification comme un tout, de la même façon qu'un niveau de dimension. Cela implique que dans l'OLAP Géographique, une mesure peut être un objet géographique aussi bien qu'un objet complexe. De plus, puisqu'une mesure peut appartenir à une ou plusieurs hiérarchies, alors dans l'OLAP Géographique, les concepts de dimension et mesure sont complètement symétriques. En utilisant cette modélisation symétrique des mesures et de dimensions, nous envisageons alors, un opérateur qui change la mesure avec un niveau d'une dimension et qui introduit les hiérarchies associées à l'ancienne mesure comme dimension d'analyse. En fait, si l'on reprend l'exemple de la Figure 4.10, grâce à cet opérateur, il serait possible de changer la mesure géographique avec les polluants et donc de transformer la mesure géographique en dimension géographique. L'application résultante présentera une mesure qui est l'objet complexe représentant les polluants et une dimension géographique « Lagoon » (Figure 4.13).

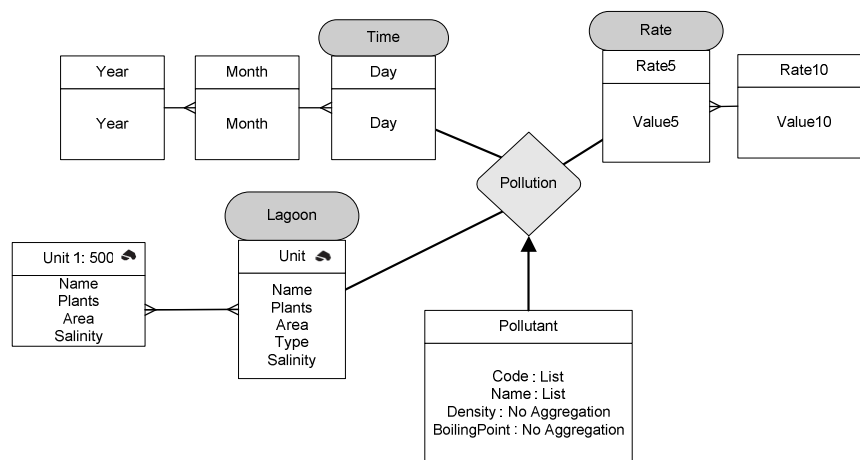


Figure 4.13. Permutation mesure et dimension : de la mesure géographique à la dimension géographique.

#### 4.3.3.3 Navigation dans la hiérarchie de la mesure

Dans notre approche une mesure géographique est liée à d'autres objets géographiques et/ou complexes à travers des hiérarchies. Par conséquent, on peut penser utiliser ces liens pour changer la granularité de la mesure, pendant le processus de navigation multidimensionnel. Nous envisageons alors des opérateurs qui remplacent les mesures avec les objets géographiques d'un niveau moins détaillé des hiérarchies des mesures. Par exemple, grâce à ces opérateurs on pourrait changer l'échelle de la mesure géographique « Unit » afin d'avoir un aperçu généralisé du phénomène géographique, et passer du modèle conceptuel représenté en Figure 4.10 au modèle de Figure 4.11.

#### 4.3.3.4 Problématiques

Les techniques de pré-agrégation des données sont fondamentales dans les systèmes OLAP pour garantir des temps de réponses satisfaisants. Les opérateurs qui modifient la structure de l'hypercube créent de nouveaux membres pour lesquels les mesures ne sont pas pré-calculées. Par conséquent, la définition des méthodes pour optimiser le calcul des mesures pour ces nouveaux membres constitue un enjeu majeur. De plus, pour le calcul de ces mesures, les connaissances de l'utilisateur par rapport au phénomène spatial et au contexte de l'application sont fondamentales, par exemple dans l'application de la lagune la distribution des polluants peut n'est pas être considérée comme une simple distribution spatiale uniforme. Cela implique que l'outil décisionnel doit fournir un langage ou une interface particulière pour permettre aux utilisateurs non informaticiens de formuler leurs modèles de recalcul au sein des requêtes multidimensionnelles.

Les opérateurs de navigation dans la hiérarchie de la mesure posent aussi différentes problématiques. Dans les modèles multidimensionnels (Pedersen et al., 2001), (Damiani et Spaccapietra, 2006) qui représentent les faits à différentes granularités, l'association entre les mesures détaillées et la mesure moins détaillée est faite a priori lors de la conception de l'application. Ils ne laissent pas la possibilité à l'utilisateur de décider la sémantique sous-jacente à l'opération de remplacement. Par exemple, on peut penser remplacer un ensemble de mesures détaillées avec une mesure moins détaillée si toutes les mesures détaillées sont présentes dans la table des faits, ou si au moins une mesure est présente dans la table des faits. Dans une vision symétrique des mesures et des dimensions, les opérations de coupe peuvent porter sur les membres de dimensions comme sur les mesures. Cela implique que les données à prendre en considération lors du changement de la granularité de la mesure peuvent être modifiées à tout moment dans le processus d'analyse spatio-multidimensionnel ou qu'une opération de changement de la mesure peut être couplée à une opération de coupe qui utilise les liens hiérarchiques des mesures.

## 4.4 Conclusion

Dans ce chapitre nous avons analysé comment les modèles SOLAP existant prennent peu en compte les aspects sémantiques de l'information géographique et la flexibilité de l'analyse spatiale.

A partir de cette analyse, nous avons formulé le concept d'OLAP Géographique, qui étend les capacités d'analyse de l'OLAP Spatial. En utilisant un cas d'étude concernant la pollution des eaux de la lagune de Venise, nous avons présenté les concepts de l'OLAP Géographique mesures et dimensions géographiques et opérateurs multidimensionnels.

La mesure géographique étend le concept de mesure spatiale aux attributs alphanumériques d'un objet géographique, qui, comme nous l'avons montré, sont nécessaires et complémentaires au processus d'analyse. De plus, une mesure peut appartenir à plusieurs hiérarchies.

Une dimension géographique est décrite par trois différents types de hiérarchies : descriptive, spatiale et de généralisation. Ces trois types de hiérarchies reflètent la sémantique des relations entre les membres de niveaux différentes. Notre approche se

focalise sur la sémantique de la hiérarchie, contrairement aux solutions SOLAP existantes, car à chaque type de hiérarchie correspond une analyse multidimensionnelle différente, qui se traduit en différentes politiques d'agrégation et de navigation.

Ensuite, nous avons montré comment dans les modèles spatio-multidimensionnels classiques la rigidité des schémas et des instances de hiérarchies empêche l'utilisation des méthodes d'analyse spatiale, et comment l'asymétrie entre dimension et mesure spatiale limite les capacités d'analyse. Nous avons alors proposé de nouveaux opérateurs spatio-multidimensionnels.

Si d'un côté tous ces nouveaux concepts permettent d'enrichir les possibilités d'analyse du SOLAP classique, de l'autre ils soulèvent de nombreuses problématiques concernant la modélisation et l'implémentation, comme cela a été détaillé dans ce chapitre. Les solutions SOLAP existantes, théoriques et pratiques, n'apparaissent pas satisfaisantes pour l'OLAP géographique. Dans les chapitres suivants nous présentons ainsi un modèle formel et un système pour OLAP Géographique.

# *Chapitre 5 : GeoCube, un modèle multidimensionnel OLAP Géographique*

Au chapitre 4, nous avons introduit le concept d'OLAP Géographique qui reformule les concepts de l'OLAP Spatial : mesures, dimensions et opérateurs. Nous avons défini les concepts de dimension et mesure géographique et nous avons introduit, en plus des opérateurs classiques de navigation, trois nouveaux types d'opérateurs : les opérateurs qui modifient dynamiquement la structure de l'hypercube, les opérateurs qui permettent de naviguer dans la hiérarchie de la mesure et l'opérateur qui change dynamiquement la mesure avec une dimension.

Dans ce chapitre, nous présentons notre modèle formel (GeoCube) et l'algèbre associée (Bimonte et al., 2005), (Bimonte et al., 2006) correspondant aux concepts introduits au chapitre 4. Notre solution est décrite dans le contexte de l'application concernant la pollution de la lagune de Venise que nous avons présentée au chapitre 4.

Le modèle de données de GeoCube est présenté dans les sections 5.1 et 5.2. L'algèbre est décrite dans la section 5.3. Nous concluons le chapitre avec un bilan comparatif entre notre solution et les modèles multidimensionnels SOLAP décrits au chapitre 3.

## **5.1 Les données multidimensionnelles**

Dans cette section, nous introduisons les concepts de GeoCube pour la représentation des données multidimensionnelles : Entité, Hiérarchie et Cube de Base.

Les concepts de Schéma d'Entité et d'Instance d'Entité représentent toutes les données de l'univers d'analyse multidimensionnelle. Ces données sont organisées en hiérarchies grâce aux concepts de Schéma de Hiérarchie et d'Instance de Hiérarchie. Le schéma de l'application multidimensionnelle et le cuboïde de base sont modélisés grâce aux concepts de Schéma de Cube de Base et d'Instance de Cube de Base.

### **5.1.1 Entité**

GeoCube représente toutes les données de l'application multidimensionnelle, membres de dimension et mesures, à travers les concepts de Schéma et d'Instance d'Entité. Un Schéma d'Entité définit la structure d'un objet complexe ou d'un objet géographique, à travers un ensemble d'attributs, par exemple un département est décrit par un nom, une

population, etc. Une Instance d'Entité représente un objet complexe ou un objet géographique. Une Instance d'Entité est décrite par un ensemble de valeurs associées aux différents attributs du Schéma d'Entité, par exemple une instance d'un département est le département du Rhône, avec une population de 1654052 habitants, etc. Pour les objets complexes, ces attributs peuvent être alphanumériques et/ou numériques. Les objets géographiques présentent en plus un attribut spatial. Les objets complexes et géographiques peuvent éventuellement avoir des attributs calculés en utilisant les autres attributs. Ces derniers correspondent à des attributs dérivés.

Nous définissons les concepts d'Attribut, et de Schéma et d'Instance d'Entité de la façon suivante :

***Définition 1. Attribut***

*Soient  $dom_1, \dots, dom_n$  des domaines de valeurs, un Attribut  $a_i$  est un identifiant alphanumérique défini sur un domaine  $dom_i = dom(a_i)$*

***Définition 2. Valeur d'Attribut***

*La valeur d'un Attribut  $a_i$ , noté  $val(a_i)$ , est une valeur qui appartient au domaine  $dom(a_i)$*

**Exemple 1.**

Un exemple d'Attribut est l'attribut population qui représente la population d'un département. Son domaine est  $dom(population) = \mathbb{N}$ .

Une valeur de l'Attribut population est  $val(population) = 254000$ .

***Définition 3. Schéma d'Entité***

*Un Schéma d'Entité  $S_e$  est un tuple  $\langle a_1, \dots, a_n, [F] \rangle$  où  $\forall i \in [1, \dots, n]$ ,  $a_i$  est un Attribut défini sur un domaine  $dom(a_i)$  et  $F$ , s'il existe, est un tuple  $\langle f_1, f_2, \dots, f_m \rangle$  où  $\forall l \in [1, \dots, m]$ ,  $f_l$  est une fonction définie sur un sous-ensemble des attributs  $a_1, \dots, a_n$*



**Définition 4. Schéma d'Entité Géographique**

Un Schéma d'Entité  $S_e$  est dit Schéma d'Entité Géographique si le domaine d'un des attributs est l'espace euclidien en deux dimensions  $\mathcal{R}^2$ . Cet attribut est appelé « attribut géométrique »

Soit  $S_e$  un Schéma d'Entité, nous notons  $\mathcal{A}(S_e)$  l'ensemble des Attributs  $\{a_1, \dots, a_n\}$ ,  $S_e.a_i$  le  $i^{\text{ème}}$  Attribut, et  $S_e.f_i$  la  $i^{\text{ème}}$  fonction.

**Exemple 2.**

Une unité de la lagune de Venise est caractérisée par son nom, le type d'activité de l'unité (industrielle, marais, etc.), une liste des plantes qui y poussent majoritairement, la salinité et la surface.

Le Schéma d'Entité Géographique qui représente une unité est  $S_{\text{unit}} = \langle \text{name\_unit}, \text{geometry}, \text{salinity}, \text{listofplants}, \text{type}, \langle f^{\text{area}} \rangle \rangle$  où :

- $\text{dom}(\text{name\_unit}) = \{ \text{'Mazzorbo'}, \text{'Murano'}, \dots \}$ ,
- $\text{geometry}$  est l'attribut géométrique,
- $\text{dom}(\text{salinity}) = \mathbb{N}$ ,
- $\text{dom}(\text{listofplants}) = \text{liste de } \{ \text{'Spartima Marittima'}, \text{'Fullica'}, \dots \}$ ,
- $\text{dom}(\text{type}) = \{ \text{'Commercial'}, \text{'Industrial'}, \dots \}$ ,
- $f^{\text{area}}: \text{dom}(\text{geometry}) \rightarrow \mathbb{R}$ .

$f^{\text{area}}$  est la fonction utilisée pour calculer la surface en utilisant la géométrie

$\mathcal{A}(S_{\text{unit}}) = \{ \text{name\_unit}, \text{geometry}, \text{salinity}, \text{listofplants}, \text{type} \}$  et  $S_{\text{unit}.a_2} = \text{geometry}$ .

La superficie est une valeur calculée à partir de la géométrie. Il s'agit d'un attribut calculé directement à partir de la composante géométrique de l'objet géographique « Unit ».

Le Schéma d'Entité qui représente un polluant est  $S_{\text{pollutant}} = \langle \text{pollutant\_name}, \text{code}, \text{density}, \text{boilingpoint} \rangle$  où :

- $\text{dom}(\text{pollutant\_name}) = \{ \text{'Ethane'}, \text{'Zinc'}, \dots \}$ ,
- $\text{dom}(\text{code}) = \{ \text{'C2H6'}, \text{'Zn'}, \dots \}$ ,

- $\text{dom}(\text{density}) = \mathbb{R}$ ,
- $\text{dom}(\text{boilingpoint}) = \mathbb{R}$ .

**Définition 5. Instance d'Entité**

Une Instance d'un Schéma d'Entité  $S_e$  est un tuple  $t_i$  tel que

- si  $F$  existe,  $t_i = \langle \text{val}(a_1), \text{val}(a_n), \text{val}(b_1), \dots, \text{val}(b_m) \rangle$  où  $\forall i \in [1, \dots, n]$ ,  $\text{val}(a_i) \in \text{dom}(a_i)$  et  $\forall j \in [1, \dots, m]$ ,  $\text{val}(b_j) = f_j(\text{val}(a_r), \dots, \text{val}(a_k))$  où  $r$  et  $k \in [1, \dots, n]$
- si  $F$  n'existe pas,  $t_i = \langle \text{val}(a_1), \dots, \text{val}(a_n) \rangle$  où  $\forall i \in [1, \dots, n]$ ,  $\text{val}(a_i) \in \text{dom}(a_i)$

Soit  $t_i$  une instance du Schéma d'Entité  $S_e$ , nous notons  $t_i.a_j$  la valeur du  $j$ -ième attribut de  $t_i$ ,  $t_i.f_j$  la valeur de l'attribut calculée par la  $j$ -ième fonction de  $t_i$  et  $I(S_e)$  l'ensemble de tuples instances de  $S_e$ .

**Exemple 3.**

Un exemple d'une instance de  $S_{\text{unit}}$  est  $t_1 = \langle \text{'Canal Fondello'}, P1, 24, \text{list}(\text{'Spartima Marittima'}, \text{'Fullica'}), \text{'Commercial'}, 245100 \rangle$  où  $P1$  est un objet spatial et  $t_{1.f}^{\text{area}} = 245100$ .

Un exemple d'une instance de  $S_{\text{pollutant}}$  est  $t_2 = \langle \text{'Zinc'}, \text{'Zn'}, 7.14, 1180 \rangle$  où  $t_2.\text{name} = \text{'Zinc'}$ .

En conclusion, le concept d'Entité (Schéma et Instance) permet de représenter les objets complexes et les objets géographiques, car une Entité représente un ensemble d'attributs/valeurs, et éventuellement, des attributs/valeurs dérivés, qui sont modélisés grâce à un ensemble de fonctions, ayant une signification comme un tout. Le concept d'Entité est utilisé dans GeoCube pour représenter sans distinction mesures et membres de dimension. Quand l'Entité représente un membre de dimension, cet ensemble d'attributs nous permet de représenter les attributs des dimensions, par exemple l'attribut représentant les plantes pour les unités de la lagune ou la densité pour les polluants. Quand une Entité représente une mesure, l'ensemble des attributs permet de modéliser un ensemble de mesures au sens classique, par exemple le nom, la géométrie, les plantes, la salinité qui identifie une unité. En outre, l'ensemble des fonctions permet de représenter les mesures ou les attributs de dimension dérivés au sens classique.

**5.1.2 Hiérarchie**

Les Entités sont organisées en hiérarchies grâce aux concepts de Schéma et d'Instance de Hiérarchie. Un niveau d'une hiérarchie est un Schéma d'Entité, et un membre est une Instance d'Entité. Le Schéma de Hiérarchie organise les niveaux d'une hiérarchie dans un treillis.

Grâce au concept d'Instance de Hiérarchie, les membres d'une hiérarchie sont organisés dans une structure à arbre. La racine est une instance du Schéma d'Entité qui est l'extrémité supérieure du treillis représenté par le Schéma de Hiérarchie. Les feuilles sont des instances du Schéma d'Entité qui est l'extrémité inférieure du treillis représenté par le Schéma de Hiérarchie.

Nous fournissons une représentation graphique, explicitée dans le paragraphe 5.1.2.2, pour chacun de ces concepts.

### 5.1.2.1 Définitions

Nous définissons ci-dessous les concepts Schéma et d'Instance de Hiérarchie.

#### **Définition 6. Schéma de Hiérarchie**

Un Schéma de Hiérarchie est un tuple  $H_h = \langle \mathcal{L}_h, \perp_h, \top_h, \#_h \rangle$  où :

- $\mathcal{L}_h$  est un ensemble de Schémas d'Entité,
- $\perp_h$  et  $\top_h$  sont deux Schémas d'Entité et  $\top_h$  contient une seule instance ('all'),
- $\#_h$  est un ordre partiel défini sur l'ensemble des niveaux de la hiérarchie  $\mathcal{L}_h \cup \perp_h \cup \top_h$  et  $\#_h$  forme un treillis où  $\perp_h$  et  $\top_h$  sont respectivement l'extrémité inférieure et l'extrémité supérieure de l'ordre.

Nous appelons niveau bottom le Schéma d'Entité  $\perp_h$ , niveau top le Schéma d'Entité  $\top_h$  et Schémas d'Entité Niveaux les Schémas d'Entité appartenant à l'ensemble  $\mathcal{L}_h \cup \perp_h \cup \top_h$ . Nous notons  $\mathcal{L}(H_h)$  les Schémas d'Entité Niveaux.

Deux Schémas de Hiérarchie  $H_i$  et  $H_j$  sont disjoints si les deux ensembles contenant leurs Schémas d'Entité Niveaux sont disjoints.

#### **Exemple 4.**

Nous détaillons dans cet exemple les Schémas de Hiérarchie présentés au chapitre 4.

Le Schéma de Hiérarchie qui représente la hiérarchie spatiale de la lagune est  $H_{\text{lagoon\_spatial}} = \langle \mathcal{L}_{\text{lagoon\_spatial}}, S_{\text{unit}}, S_{\text{all\_unit}}, \#_{\text{lagoon\_spatial}} \rangle$  où  $\mathcal{L}_{\text{lagoon\_spatial}} = \{S_{\text{zone}}\}$  et  $(S_{\text{unit}} \#_{\text{lagoon\_spatial}} S_{\text{zone}})$  et  $(S_{\text{zone}} \#_{\text{lagoon\_spatial}} S_{\text{all\_unit}})$ . Le niveau top de  $H_{\text{lagoon\_spatial}}$  est  $S_{\text{all\_unit}}$

Le Schéma de Hiérarchie qui représente la hiérarchie de généralisation de la lagune est  $H_{\text{lagoon\_generalization}} = \langle \mathcal{L}_{\text{lagoon\_generalization}}, S_{\text{unit}}, S_{\text{all\_unit}}, \#_{\text{lagoon\_generalization}} \rangle$  où  $\mathcal{L}_{\text{lagoon\_generalization}} = \{S_{\text{unit}1:500}\}$  et  $(S_{\text{unit}} \#_{\text{lagoon\_generalization}} S_{\text{unit}1:500})$ ,  $(S_{\text{unit}1:500} \#_{\text{lagoon\_generalization}} S_{\text{all\_unit}})$  et  $(S_{\text{unit}1:500} \#_{\text{lagoon\_generalization}} S_{\text{all\_unit}})$ . Cette

hiérarchie représente les unités à l'échelle 1:1000 ( $S_{unit}$ ) et à l'échelle moins détaillée 1:500 ( $S_{unit1:500}$ ).

$H_{lagoon\_descriptive}$  est une hiérarchie qui classe les unités par rapport à leur type.  $H_{lagoon\_descriptive} = \langle \mathcal{L}_{lagoon\_descriptive}, S_{unit}, S_{all\_unit}, \#_{lagoon\_descriptive} \rangle$  où  $\mathcal{L}_{lagoon\_descriptive} = \{S_{type\_unit}\}$  et  $(S_{unit} \#_{lagoon\_descriptive} S_{type\_unit}), (S_{type\_unit} \#_{type} S_{all\_unit})$ .

$H_{time} = \langle \mathcal{L}_{time}, S_{day}, S_{all\_day}, \#_{time} \rangle$  où  $\mathcal{L}_{time} = \{S_{month}, S_{year}\}$  et  $(S_{day} \#_{time} S_{month}), (S_{month} \#_{time} S_{year}), (S_{year} \#_{time} S_{all\_day})$ . Les Schémas d'Entité Niveaux de  $H_{time}$  sont  $S_{day}, S_{month}, S_{year}$  et  $S_{all\_day}$ .  $H_{time}$  représente la dimension temporelle.

Le Schéma de Hiérarchie qui représente la hiérarchie des polluants est :  $H_{pollutants} = \langle \mathcal{L}_{pollutants}, S_{pollutant}, S_{all\_pollutant}, \#_{pollutants} \rangle$  où  $\mathcal{L}_{pollutants} = \{S_{boundsType}, S_{nbCarbonsAtoms}, S_{type}\}, (S_{pollutant} \#_{pollutants} S_{boundsType}), (S_{boundsType} \#_{pollutants} S_{nbCarbonsAtoms}), (S_{nbCarbonsAtoms} \#_{pollutants} S_{type}), (S_{type} \#_{pollutants} S_{all\_pollutant})$  et  $(S_{pollutant} \#_{pollutants} S_{nbCarbonsAtoms})$ .

Le Schéma de Hiérarchie qui représente la hiérarchie des valeurs de la pollution est  $H_{rate} = \langle \mathcal{L}_{rate}, S_{rate5}, S_{all\_rate}, \#_{rate} \rangle$  où  $\mathcal{L}_{rate} = \{S_{rate10}\}$  et  $(S_{rate5} \#_{rate} S_{rate10}), (S_{rate10} \#_{rate} S_{all\_rate})$ .  $S_{rate5}$  et  $S_{rate10}$  représentent une classification des valeurs de pollution par intervalles de 5 mg/l et 10 mg/l, par exemple [0, 4] mg/l pour  $S_{rate5}$  et [20, 29] mg/l pour  $S_{rate10}$ . Le niveau bottom de  $H_{rate}$  est  $S_{rate5}$ .

### Exemple 5.

Les Schémas de Hiérarchie  $H_{lagoon\_spatial}$  et  $H_{pollutants}$  sont disjoints.

#### **Définition 7. Instance de Hiérarchie**

Une Instance d'un Schéma de Hiérarchie  $H_h$  est un ordre partiel  $\uparrow_h$  défini sur  $\mathcal{L}(H_h)$ , tel que

- si  $t_i \uparrow_h t_j$  alors  $S_i \#_h S_j$ , où  $t_i \in I(S_i)$  et  $t_j \in I(S_j)$
- $\forall t_i \in I(\mathcal{L}(H_h)) / I(\sqrt{h}), \exists t_j \in I(\mathcal{L}(H_h))$  tel que  $t_j \uparrow_h t_i$
- $\forall t_i \in I(\mathcal{L}(H_h)) / I(\mathcal{L}_h), \exists t_j \in I(\mathcal{L}(H_h))$  tel que  $t_i \uparrow_h t_j$

Nous notons  $leafs(H_h, t_i)$  l'ensemble des feuilles de l'arbre formé par  $\uparrow_h$  qui a comme racine  $t_i$ .

### Exemple 6.

Pour simplifier la lecture, nous utilisons seulement les valeurs de l'attribut *name\_unit* pour identifier une instance de  $S_{unit}$  et les valeurs de l'attribut *name\_zone*

pour identifier une instance de  $S_{zone}$ . L'instance du Schéma de Hiérarchie  $H_{lagoon\_spatial}$  est :

(Coa di Latte  $\uparrow_{lagoon\_spatial}$  Bocca Lido), (Canal Fondello  $\uparrow_{lagoon\_spatial}$  Bocca Lido), (Mazzorbo  $\uparrow_{lagoon\_spatial}$  North Swam), (Murano  $\uparrow_{lagoon\_spatial}$  North Swam), (Chioggia  $\uparrow_{lagoon\_spatial}$  Bocca Chioggia), (Romea  $\uparrow_{lagoon\_spatial}$  Bocca Chioggia), (Ronzei  $\uparrow_{lagoon\_spatial}$  South Swam), (Tessera  $\uparrow_{lagoon\_spatial}$  South Swam), (South Swam  $\uparrow_{lagoon\_spatial}$  All\_units), (Bocca Chioggia  $\uparrow_{lagoon\_spatial}$  All\_units), (North Swam  $\uparrow_{lagoon\_spatial}$  All\_units) et (Bocca Lido  $\uparrow_{lagoon\_spatial}$  All\_units).

$leaves(H_{lagoon\_spatial}, Bocca Chioggia) = \{Romea, Chioggia\}$ .

Un exemple de l'instance de la hiérarchie  $H_{lagoon\_descriptive}$  est :

(Coa di Latte  $\uparrow_{lagoon\_descriptive}$  Industrial), (Murano  $\uparrow_{lagoon\_descriptive}$  Commercial), (Murano  $\uparrow_{lagoon\_descriptive}$  Industrial), (Industrial  $\uparrow_{lagoon\_descriptive}$  All\_units) et (Commercial  $\uparrow_{lagoon\_descriptive}$  All\_units).

### 5.1.2.2 Représentation graphique

Nous présentons dans ce paragraphe une représentation graphique des concepts de Schéma et d'Instance de Hiérarchie.

Un Schéma d'Entité  $S_e$  est représenté par la notation montrée en Figure 5.1a. L'ordre partiel  $\#_h$  est représenté par la notation montrée en Figure 5.1b où la flèche de  $S_i$  à  $S_j$  signifie que  $S_i \#_h S_j$ . L'ordre partiel  $\uparrow_h$  est représenté par la notation montrée en Figure 5.1c, où une flèche de  $t_i$  à  $t_j$  signifie que  $t_i \uparrow_h t_j$ .

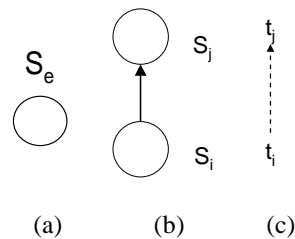


Figure 5.1. Représentation graphique a) Schéma d'Entité b) Ordre partiel des Schémas d'Entité Niveaux  $\#_h$  c) Ordre partiel des instances des Schémas d'Entité Niveaux  $\uparrow_h$ .

Nous appelons graphe de navigation la notation graphique utilisée pour représenter un Schéma de Hiérarchie.

#### Exemple 7.

Le graphe de navigation de la hiérarchie des polluants  $H_{pollutants}$  est montré en Figure 5.2a. L'instance de la hiérarchie est représentée en Figure 5.2b.

La représentation graphique de la hiérarchie spatiale de la lagune  $H_{lagoon\_spatial}$  (Schéma et Instance) est illustrée en Figure 5.3.

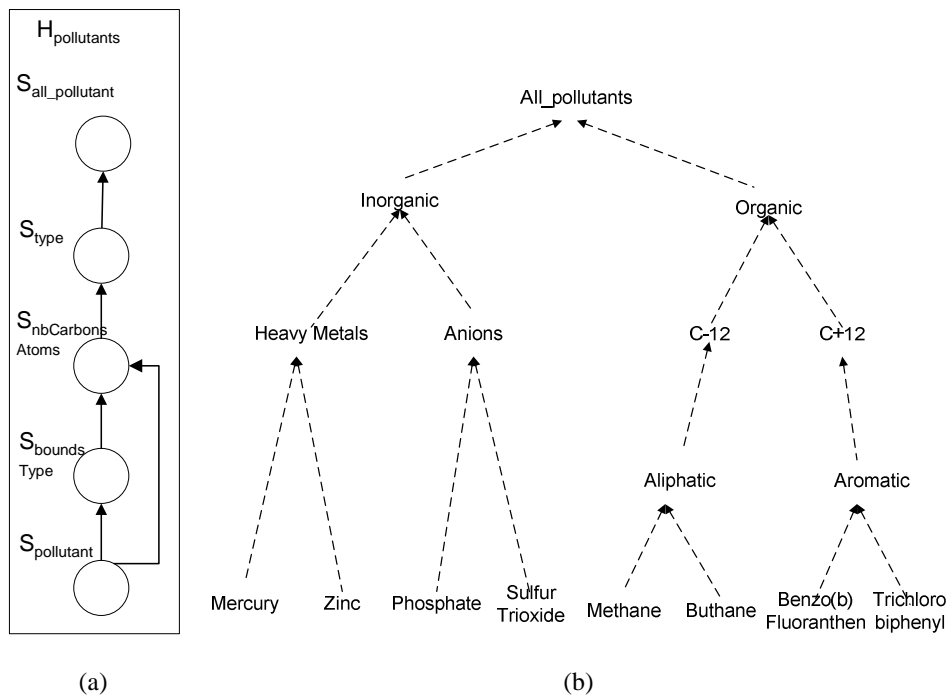


Figure 5.2. Hiérarchie de polluants a) Schéma b) Instance.

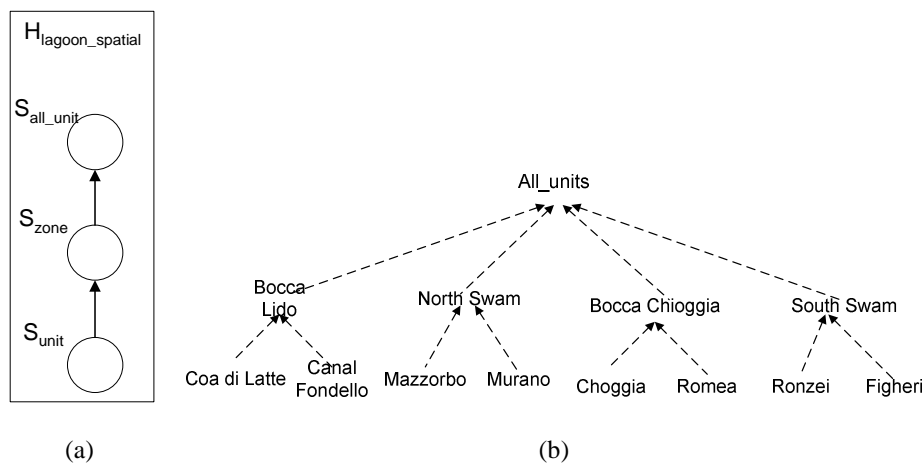


Figure 5.3. Hiérarchie spatiale de la lagune a) Schéma b) Instance.

Les concepts de Schéma et d'Instance de Hiérarchie nous permettent de modéliser les hiérarchies complexes non-couvrantes et non-strictes. Un exemple de hiérarchie non-couvrante est la hiérarchie des polluants où par exemple les membres « Zinc » et « Mercury » sautent les niveaux représentés par le type de lien atomique. Un exemple de hiérarchie non-strictes est la hiérarchie descriptive  $H_{lagoon\_descriptive}$ . Dans cette hiérarchie l'unité Murano est classée zone industrielle et zone commerciale.

Ces exemples montrent dans quelle mesure la prise en compte des hiérarchies complexes est nécessaire dans l'OLAP Géographique. Ce besoin est évident aussi pour la

représentation des hiérarchies de généralisation, comme nous l'avons souligné au chapitre 4. En effet, le processus de généralisation crée une structure hiérarchique complexe où les objets géographiques peuvent être liés par des relations  $n : 1$ ,  $n : m$  et  $1 : 0$  (cf chap 2 § 1.3). En utilisant une terminologie OLAP, une hiérarchie de généralisation peut être stricte ( $n : 1$ ) et non-strictes ( $n : m$ ). La relation  $1 : 0$  peut être modélisée en utilisant une hiérarchie non couvrante, où les membres n'ayant pas de pères « sautent » directement au membre du niveau top de la hiérarchie.

Le concept de Hiérarchie (Schéma et Instance) nous permet ainsi d'organiser les données en hiérarchies complexes, grâce aux ordres partiels définis par le Schéma et l'Instance de Hiérarchie.

### 5.1.3 Cube de Base

Le Schéma de Cube de Base représente le schéma conceptuel de l'application multidimensionnelle et l'Instance de Cube de Base représente le cuboïde de base (cf chap 1 § 3.2.1), soit la table de faits où toutes les dimensions sont aux niveaux les plus détaillés.

Au chapitre 4, nous avons souligné l'intérêt d'une modélisation complètement symétrique des mesures et des dimensions dans le processus d'analyse spatio-multidimensionnelle. Cette symétrie se traduit par la représentation des membres des dimensions et des mesures de la même façon, et par l'appartenance des mesures à des hiérarchies.

Nous définissons alors les concepts de Schéma et Instance de Cube de Base de la manière suivante :

#### ***Définition 8. Schéma de Cube de Base***

*Un Schéma de Cube de Base  $BC_{bc}$  est un tuple  $\langle H_1, \dots, H_{m+1}, \delta \rangle$  où :*

- $\forall i \in [1, \dots, m+1]$ ,  $H_i$  est un Schéma de Hiérarchie et toutes les hiérarchies  $H_i$  sont disjointes
- $\delta$  est une fonction booléenne définie sur  $I(L_1) \times \dots \times I(L_{m+1})$  ( $\delta$  est définie sur les niveaux bottom des hiérarchies  $H_1, \dots, H_{m+1}$ )

Nous notons  $H(BC_{bc}, S_e)$  le Schéma de Hiérarchie du Cube de Base  $BC_{bc}$  auquel appartient le Schéma d'Entité  $S_e$ .

Le Schéma de Cube de Base est un ensemble de Schémas de Hiérarchie et une fonction booléenne. Cette fonction peut être représentée par un espace multidimensionnel où les instances des membres des niveaux les plus détaillés des dimensions sont projetées sur les axes et les points sont la valeur vraie ou fausse (Figure 5.4). La valeur « vrai » signifie que les données correspondant à ces coordonnées particulières d'instances des Schémas d'Entité sont présentes dans la base de données multidimensionnelle. Grâce à cette

modélisation de l'espace multidimensionnel, les Schémas de Hiérarchie du Cube de Base jouent tous le rôle de dimension. Tous leurs niveaux les plus détaillés peuvent être utilisés comme mesures. Une mesure peut donc appartenir à une hiérarchie de la même façon qu'un niveau de dimension. L'application multidimensionnelle modélisée par le Schéma de Cube de Base ne fait alors aucune distinction entre mesure et dimension, autorisant une vision symétrique des dimensions et des mesures.

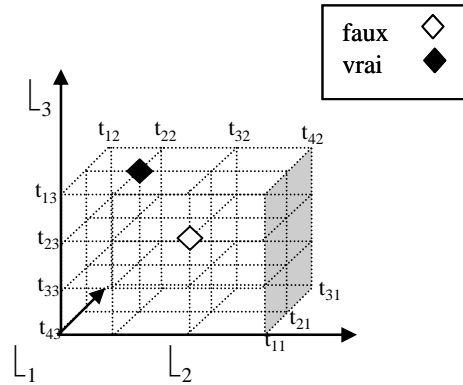


Figure 5.4. Représentation de la fonction booléenne d'un Cube de Base avec trois Schémas de Hiérarchie.

**Exemple 8.**

Le Schéma de Cube de Base de l'application concernant la pollution des eaux de la lagune de Venise est  $BC_{corila} = \langle H_{pollutants}, H_{time}, H_{lagoon\_spatial}, H_{rate}, \delta \rangle$  où  $\delta$  est définie sur  $I(S_{pollutant}) \times I(S_{day}) \times I(S_{unit}) \times I(S_{rate5})$ .

**Définition 9. Instance de Cube de Base**

Une Instance d'un Schéma de Cube de Base  $BC_{bc}$ , noté  $I(BC_{bc})$ , est un ensemble de tuples tel que  $I(BC_{bc}) = \{ \langle t_1^j, \dots, t_{m+1}^j \rangle, j=1, \dots, p \}$  où  $\forall i \in [1, \dots, m+1], t_i^j \in I(L_i)$  et  $\delta(t_1^j, \dots, t_{m+1}^j)$  est vrai

L'ensemble des tuples qui représente une instance d'un Schéma de Cube de Base correspond aux combinaisons des instances de membres présents dans la base de données multidimensionnelle. Pour ces tuples la fonction booléenne est vraie. Puisque tous les membres appartiennent aux niveaux les plus détaillés, une Instance de Cube de Base représente une table de faits sans faits.



**Exemple 9.**

L'instance du Schéma de Cube de Base  $BC_{corila}$  est :

$\langle\langle$ 'Zinc', 'Zn',7.14,1180), <19-9-05), <'Mazzorbo', P1, 24, list('Spartima Marittima', 'Fullica'), 'Commercial', 245100) , <45-49)>\rangle, \langle\langle'Methane', 'CH14', 7.17, 111.55), <19-9-05), <'Mazzorbo', P1, 24, list('Spartima Marittima', 'Fullica'), Commercial, 245100), <'30-34')>\rangle, etc.

La Figure 5.5 montre un exemple des valeurs de la fonction booléenne  $\delta$  de  $BC_{corila}$ . L'ensemble de tous les tuples de  $I(BC_{corila})$  est montré dans le Tableau 5.1. Par souci de simplicité, nous représentons les instances des Schémas d'Entité en utilisant seulement une valeur d'un attribut.

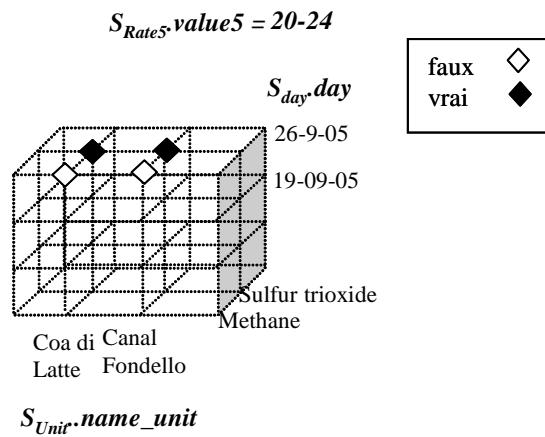


Figure 5.5. Représentation de la fonction booléenne du Cube de Base  $BC_{corila}$ .

Grâce à cette structuration particulière de l'espace multidimensionnel, les données associées à l'instance d'un Schéma de Cube de Base peuvent être utilisées pour répondre à des requêtes multidimensionnelles qui utilisent n'importe quel niveau le plus détaillé des Schémas de Hiérarchie comme mesure, et les autres Schémas de Hiérarchie comme axes d'analyse. Par exemple, si l'on considère l'exemple 9, une requête qui utilise les unités comme mesures peut être : « *Quelles sont les unités polluées pour chaque polluant, pour chaque jour et pour chaque valeur de pollution ?* » ; une requête qui porte sur les polluants peut être « *Quels sont les polluants présents dans chaque unité, pour chaque jour et pour chaque valeur de pollution ?* ».

De plus, cette approche autorise la modélisation des relations plusieurs à plusieurs entre faits et dimensions, et des mesures multi-valuées. Le Tableau 5.1 montre que le 19/9/05, la même unité « Mazzorbo » est polluée par deux polluants différents, « Methane » et « Sulfur Trioxide », avec une valeur de pollution de 30-34 mg/l, et que le 19/9/05, les unités « Mazzorbo » et « Murano » sont polluées par le zinc avec une valeur de pollution de 45-49 mg/l.

$S_{pollutant.name\_pollutant}$	$S_{day.day}$	$S_{unit.name\_unit}$	$S_{rate5.value5}$
Zinc	19-9-05	Mazzorbo	45-49
Sulfur Trioxide	19-9-05	Mazzorbo	30-34
Methane	19-9-05	Mazzorbo	30-34
Zinc	19-9-05	Canal Fondello	30-34
Benzo(b)fluoronthene	19-9-05	Murano	45-49
Methane	19-9-05	Murano	30-34
Zinc	19-9-05	Tessera	45-49
Methane	19-9-05	Coa di Latte	30-34
Zinc	19-9-05	Murano	45-49
Methane	26-9-05	Tessera	30-34
Sulfur Trioxide	26-9-05	Canal Fondello	20-24
Sulfur Trioxide	26-9-05	Coa di Latte	20-24

Tableau 5.1. Instance du Cube de Base BC<sub>corila</sub>.

Le concept de Cube de Base (Schéma et Instance) définit ainsi le schéma et les données d'une application multidimensionnelle sans faire de distinction entre mesure et dimension.

## 5.2 Requêtes multidimensionnelles

Dans cette section, nous définissons les concepts de Vue (Schéma et Instance) et Mode d'Agrégation pour représenter respectivement le résultat d'une requête multidimensionnelle et l'agrégation des objets complexes ou géographiques.

### 5.2.1 Mode d'Agrégation

Le concept de Mode d'Agrégation est crucial dans notre modèle car il permet d'agréger des objets complexes ou géographiques. L'idée est de considérer un Schéma d'Entité comme mesure et un autre Schéma d'Entité comme mesure agrégée, et de lier les instances des mesures détaillées à une instance de la mesure agrégée à travers des fonctions d'agrégation. Une fonction est définie pour chaque attribut de la mesure agrégée.

Les fonctions d'agrégation appliquées aux différents attributs de l'objet complexe ou géographique dépendent du type d'attribut. Nous distinguons deux types de fonctions : les fonctions topologiques et les fonctions alphanumériques. Les fonctions d'agrégation topologique (Shekar et al., 2001) (union spatiale, intersection, barycentre, convex hull, etc.) sont applicables à l'attribut spatial des objets géographiques. Les fonctions d'agrégation alphanumériques sont des fonctions que l'on peut appliquer aux attributs descriptifs. Ces fonctions d'agrégation peuvent être mathématiques (e.g. la moyenne de la salinité), ou définies par l'utilisateur (e.g. le type d'un ensemble des unités peut être défini grâce à des fonctions de logique floue). Dans le cas des mesures géographiques, ces fonctions peuvent exploiter les résultats des fonctions d'agrégation topologiques et les attributs géométriques et descriptifs des objets géographiques originaux.

Ainsi, les fonctions d'agrégation que l'on applique aux objets complexes ou géographiques ne sont pas limitées aux fonctions d'agrégation classiques comme SUM, ou AVG, mais peuvent être très complexes. De plus, certains attributs ne peuvent pas être agrégés (par exemple les noms des unités) et les objets agrégés peuvent présenter des attributs de type différent. Un objet complexe/géographique agrégé peut aussi avoir moins ou plus d'attributs que l'objet complexe/géographique représentant la mesure détaillée. Par exemple, dans une application multidimensionnelle pour le contrôle des accidents routiers une mesure peut être représentée par la localisation de l'accident, qui est un objet spatial de type point. Une agrégation spatiale possible peut être le convex hull. Le résultat de cette agrégation spatiale est un polygone. La surface de cette nouvelle géométrie devient un nouvel attribut dans la mesure agrégée qui peut être utilisée dans le processus décisionnel.

Nous définissons ainsi un Mode d'Agrégation :

**Définition 10. Mode d'Agrégation**

Un Mode d'Agrégation  $\Theta_k$  est un tuple  $\langle S_a, S_b, \Phi \rangle$  où :

- $S_a$  est un Schéma d'Entité  $\langle a_1, \dots, a_m, [F_a] \rangle$  appelé mesure détaillée
- $S_b$  est un Schéma d'Entité  $\langle b_1, \dots, b_p, [F_b] \rangle$  appelé mesure agrégée
- $\Phi$  est ensemble de  $p$  fonctions ad hoc d'agrégation  $\phi_i$ .

**Définition 11. Agrégat de**

L'agrégation de  $n$  instances  $t_1, \dots, t_n$  de  $S_a$  en utilisant un Mode d'Agrégation  $\Theta_k = \langle S_a, S_b, \Phi \rangle$  est une instance de  $t_b$  de  $S_b$ ,  $t_b = \langle val(b_1), \dots, val(b_p), val(d_1), \dots, val(d_m) \rangle$  tel que :

- $\forall j \in [1, \dots, p] \text{ val}(b_j) = \phi_j(t_i.a_1, \dots, t_i.a_k)$  avec
- $t_i \in I(S_a)$  et  $a_r \in \mathcal{A}(S_a)$  où  $r \in [1, \dots, k]$  et  $l$  et  $k \in [1, \dots, m]$
- $\forall i \in [1, \dots, l], \text{ val}(d_i) = f_i(\text{val}(b_r), \dots, \text{val}(b_k))$  où  $r, k \in [1, \dots, m]$  si  $F_b$  existe

Dans ce cas nous disons que  $t_b$  est l'agrégat de  $t_1, \dots, t_n$ .

**Exemple 10.**

Soit  $S_{\text{unit\_fusion}} = \langle \text{geometry, salinity, listofplants, type, } \{ \{ f^{\text{area}} \} \} \rangle$  où :

- geometry est l'attribut géométrique
- $\text{dom}(\text{salinity}) = \mathbb{N}$ ,
- $\text{dom}(\text{listofplants}) = \text{liste de } \{ \text{'Spartima Marittima'}, \text{'Fullica'}, \dots \}$ ,
- $\text{dom}(\text{type}) = \{ \text{'Commercial'}, \dots \}$ ,
- $f^{\text{area}} : \text{dom}(\text{geometry}) \rightarrow \mathbb{R}$ .

$f^{\text{area}}$  est la fonction utilisée pour calculer la surface en utilisant la géométrie

Alors, un exemple de Mode d'Agrégation représentant l'agrégation des unités de la lagune est (Figure 5.6) :  $\Theta_{\text{fusion-unit}} = \langle S_{\text{unit}}, S_{\text{unit\_fusion}}, \Phi_{\text{unit\_fusion}} \rangle$  où  $\Phi_{\text{unit\_fusion}} =$

- $\phi_{\text{geometry}} : \text{dom}(S_{\text{unit}}.\text{geometry})^n \rightarrow \text{dom}(S_{\text{unit\_fusion}}.\text{geometry}) // \text{fusion}$
- $\phi_{\text{salinity}} : \text{dom}(S_{\text{unit}}.\text{salinity})^n \rightarrow \text{dom}(S_{\text{unit\_fusion}}.\text{salinity}) // \text{moyenne}$
- $\phi_{\text{listofplants}} : \text{dom}(S_{\text{unit}}.\text{listofplants})^n \rightarrow \text{dom}(S_{\text{unit\_fusion}}.\text{listofplants}) // \text{liste}$
- $\phi_{\text{type}} : (\text{dom}(S_{\text{unit}}.\text{geometry}) \times \text{dom}(S_{\text{unit}}.\text{type}))^n \rightarrow \text{dom}(S_{\text{unit\_fusion}}.\text{type}) // \text{moyenne pondérée sur la surface}$ .

Notons que le Mode d'Agrégation  $\Theta_{\text{fusion-unit}}$  ne définit pas la fonction d'agrégation pour la surface car elle est calculée en utilisant la fonction  $f^{\text{area}}$  définie par le Schéma d'Entité.

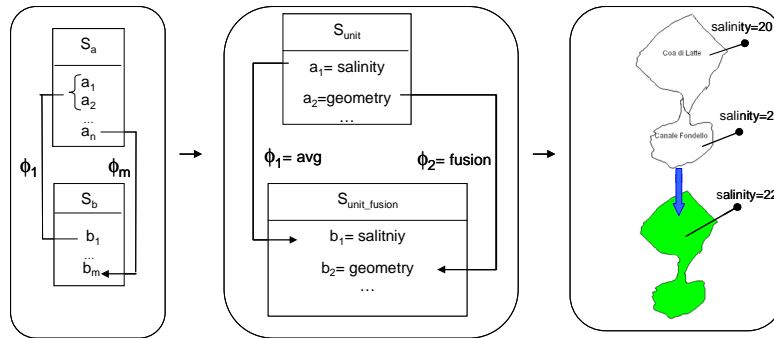


Figure 5.6. Mode d'Agrégation pour l'agrégation des objets géographiques représentant les unités de la lagune.

Le Mode d'Agrégation définit donc comment agréger des objets complexes ou géographiques. Il utilise un ensemble de fonctions d'agrégations qui lient les valeurs des attributs des Instances d'Entité représentant les mesures détaillées et les valeurs des attributs des Instances d'Entité représentant les mesures agrégées.

Nous présentons, comme pour les hiérarchies, une représentation graphique du Mode d'Agrégation. La notation graphique utilisée pour représenter un Mode d'Agrégation est montrée en Figure 5.7. Nous appelons cette représentation graphique Graphe d'Agrégation.

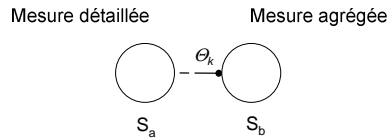


Figure 5.7. Représentation graphique du Mode d'Agrégation.

### 5.2.2 Vue

Une Vue représente le résultat d'une requête multidimensionnelle. Contrairement au Schéma de Cube de Base, le Schéma de Vue définit l'objet utilisé comme mesure, un Mode d'agrégation pour agréger cette mesure, et les niveaux des dimensions utilisées dans la requête. Une Instance de Vue représente le cuboïde dont les niveaux des dimensions sont définis par le Schéma de Vue.

Un Schéma de Vue est formé par un Schéma de Cube de Base, un Mode d'Agrégation, un ensemble de Schémas d'Entité, et une fonction booléenne :

- le Cube de Base représente le modèle conceptuel de l'application multidimensionnelle et les données de la table des faits
- le Mode d'Agrégation définit comment agréger les mesures. La mesure détaillée est le niveau le plus détaillé d'une hiérarchie du Cube de Base Schéma. Ce niveau est choisi comme étant la mesure de l'application.
- les Schémas d'Entité correspondent aux niveaux du cuboïde représenté par la Vue. Chaque Schéma d'Entité appartient à un Schéma de Hiérarchie du Cube de Base différent de celui qui contient la mesure.
- la fonction booléenne représente les données multidimensionnelles.

La fonction booléenne peut être ainsi vue comme un espace multidimensionnel où les instances des membres des niveaux des dimensions et la mesure agrégée du Mode d'Agrégation sont projetées sur les axes, et les points sont les valeurs booléennes. La valeur fautive signifie que les données correspondant à cette combinaison d'instances d'Entités ne sont pas présentes dans le cuboïde représenté par la Vue. Puisque l'instance d'une Vue où les niveaux des dimensions sont les plus détaillés correspond au cuboïde de base, la fonction booléenne de la Vue est alors utilisée pour représenter un cuboïde qui n'est pas le cuboïde de base. Les cellules de ce cuboïde contiennent les instances de la mesure agrégée. Au contraire, le cuboïde de base est représenté par une instance du Cube de Base et ses cellules contiennent les instances de la mesure détaillée.

Nous définissons le Schéma et l'Instance de Vue de la façon suivante :

**Définition 12. Schéma de Vue**

Un Schéma de Vue  $V_v$  est un tuple  $\langle BC_{bc}, \mathcal{L}, \Theta_k, \gamma \rangle$  où :

- $BC_{bc} = \langle H_{b1}, \dots, H_{bm+1}, \delta \rangle$  est un Schéma de Cube de Base,
- $\mathcal{L}$  est un tuple de Schémas d'Entité  $\langle S_{L1}, \dots, S_{Lm} \rangle$  tel que  $\forall i, j \in [1, \dots, m], \exists l$  et  $\exists p \in [1, \dots, m]$ , tels que  $S_{Li} \in L(H_{bl}), S_{Lj} \in L(H_{bp})$  et  $l \neq p$

( $\mathcal{L}$  représente l'ensemble des niveaux ( $S_{Li}$ ) des dimensions du cuboïde. Tous les Schémas d'Entité ( $S_{Li}$ ) de  $\mathcal{L}$  appartiennent à des hiérarchies différentes de  $BC_{bc}$ )

- $\Theta_k$  est un Mode d'Agrégation  $\langle S_f, S_{af}, \Phi \rangle$  tel que :
  - $\exists h \in \{H_{b1}, \dots, H_{bm+1}\}$  tel que  $S_f = L_h$
  - $\forall i \in [1, \dots, m], S_f \notin L(H(BC_{bc}, S_{Li}))$

( $S_f$  représente la mesure et il est le niveau bottom de la hiérarchie qui n'est pas utilisée pour la définition des niveaux du cuboïde ( $\mathcal{L}$ ))

- $\gamma$  est une fonction booléenne définie sur  $I(S_{L1}) \times \dots \times I(S_{Lm}) \times I(S_{af})$

Nous appelons Vue de Base un Schéma de Vue  $V_v$  où toutes les Schémas d'Entité de  $\mathcal{L}$  sont des niveaux bottom, et Niveaux de la Vue l'ensemble des Schémas d'Entité de  $\mathcal{L}$ .

**Définition 13. Instance de Vue**

Une Instance d'un Schéma de Vue  $V_v$ , noté  $I(V_v)$ , est un ensemble de tuples tel que :

- si  $V_v$  n'est pas une Vue de Base, alors  $I(V_v) = \{ \langle t_1^j, \dots, t_{m+1}^j \rangle, j=1, \dots, p \}$  où :
  - $\forall i \in [1, \dots, m], t_i^j \in I(S_{Li})$  où  $S_{Li} \in \mathcal{L}$
  - $t_{m+1}^j \in I(S_{af})$
  - $\gamma(t_1^j, \dots, t_{m+1}^j)$  est vrai
- si  $V_v$  est une Vue de Base, alors  $I(V_v) = I(BC_{bc})$

La représentation graphique d'un Schéma de Vue est composée par un ensemble de graphes de navigation et un graphe d'agrégation (Figure 5.8a). En particulier, pour chaque

hiérarchie du Cube de Base, un graphe de navigation (Figure 5.8b) est présent. Les Schémas d'Entité de graphes de navigation utilisés comme Niveaux de la Vue sont colorés en noir (Figure 5.8c). Le Schéma d'Entité du graphe d'agrégation représentant la mesure est coloré en gris (Figure 5.8d). Dans le cas où le Schéma de Vue est une Vue de Base alors ce Schéma d'Entité est la mesure détaillée du graphe d'agrégation. Autrement il est la mesure agrégée du graphe d'agrégation.

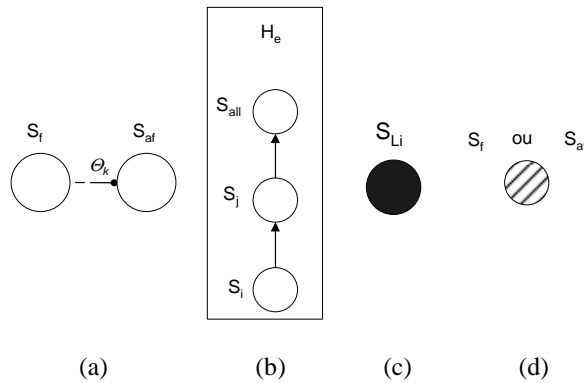


Figure 5.8. Représentation graphique du Schéma de Vue a) Graphe d'Aggrégation b) Graphe de Navigation c) Niveaux de la Vue d) Mesure détaillée ou mesure agrégée.

**Exemple 11.**

Un exemple de Vue de Base qui a comme mesure les unités de la lagune est le suivant:  $V_{\text{corila-unit}} = \langle BC_{\text{corila}}, \langle S_{\text{pollutant}}, S_{\text{day}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-unit}}, \gamma \rangle$ .

La représentation graphique de  $V_{\text{corila-unit}}$  est montrée en Figure 5.19. Son instance est égale à l'instance du Cube de Base  $BC_{\text{corila}}$  car  $V_{\text{corila-unit}}$  est une Vue de Base (Tableau 5.2).

Le Tableau 5.2 nous montre par exemple que le 19/9/05 les unités Mazzorbo, Murano et Tessera sont polluées au zinc avec une valeur de 45-49 mg/l, et que le même jour Murano était polluée au benzo(b)fluronthene avec la même valeur.

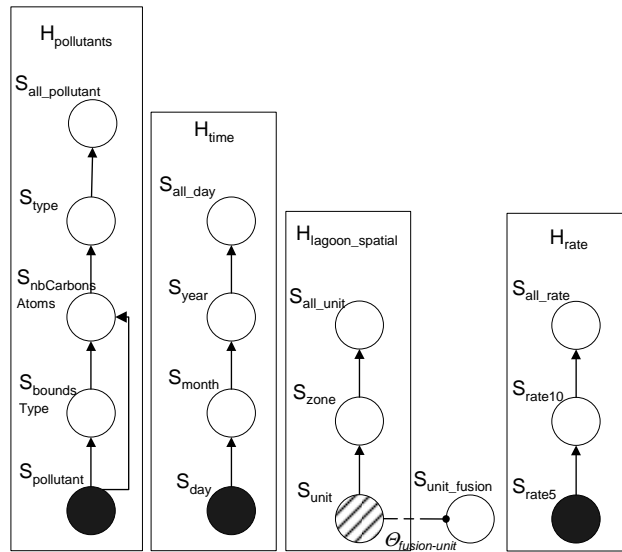


Figure 5.9. Représentation graphique du Schéma de Vue  $V_{\text{corila-unit}}$  (Vue de Base).

$S_{\text{pollutant.name\_pollutant}}$	$S_{\text{day.day}}$	$S_{\text{unit.name}}$	$S_{\text{rate5.value5}}$
Zinc	19-9-05	Mazzorbo	45-49
Sulfur Trioxide	19-9-05	Mazzorbo	30-34
Methane	19-9-05	Mazzorbo	30-34
Zinc	19-9-05	Canal Fondello	30-34
Benzo(b)fluoronthene	19-9-05	Murano	45-49
Methane	19-9-05	Murano	30-34
Zinc	19-9-05	Tessera	45-49
Methane	19-9-05	Coa di Latte	30 -34
Zinc	19-9-05	Murano	45-49
Sulfur Trioxide	26-9-05	Coa di Latte	20-24
Methane	26-9-05	Tessera	30-34
Sulfur Trioxide	26-9-05	Canal Fondello	20-24

Tableau 5.2. Instance de  $V_{\text{corila-unit}}$ .

**Exemple 12.**

Un exemple de Vue de Base qui présente les polluants comme mesures est le suivant :  $V_{\text{corila-pollutant}} = \langle BC_{\text{corila}}, \langle S_{\text{unit}}, S_{\text{day}}, S_{\text{rate5}} \rangle, \Theta_{\text{list-pollutant}}, \gamma \rangle$ . La représentation graphique ce Schéma de Vue est illustrée en Figure 5.10.



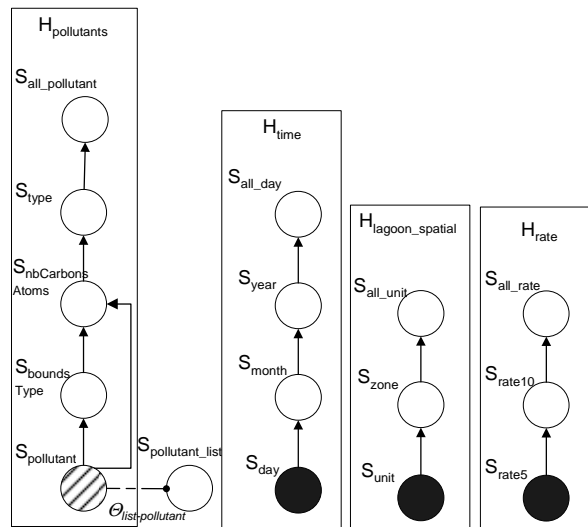


Figure 5.10. Représentation graphique du Schéma de Vue  $V_{\text{corila-pollutant}}$  (Vue de Base).

Les exemples 11 et 12 montrent que, même si  $V_{\text{corila-pollutant}}$  et  $V_{\text{corila-unit}}$  représentent deux analyses différentes, une avec les unités comme mesures et l'autre avec les polluants comme mesures, elles présentent le même Cube de Base, car elles reposent sur la même table des faits. Ce qui les distingue est la mesure, le Mode d'Agrégation et les niveaux de dimensions utilisés.

Nous montrons maintenant un exemple de Vue qui ne représente pas un cuboïde de base.

### Exemple 13.

Un exemple de Schéma de Vue avec les unités comme mesures est le suivant :  
 $V_{\text{corila-unit-month}} = \langle BC_{\text{corila}}, \langle S_{\text{pollutant}}, S_{\text{month}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-unit}}, \gamma \rangle$ .

Sa représentation graphique est montrée en Figure 5.11 et son instance dans le Tableau 5.3. Nous notons que les mesures sont les instances de la mesure agrégée du Mode d'Agrégation car  $V_{\text{corila-unit-month}}$  utilise le niveau mois ( $S_{\text{month}}$ ) de la hiérarchie temporelle ( $H_{\text{time}}$ ) qui n'est pas le niveau le plus détaillé.

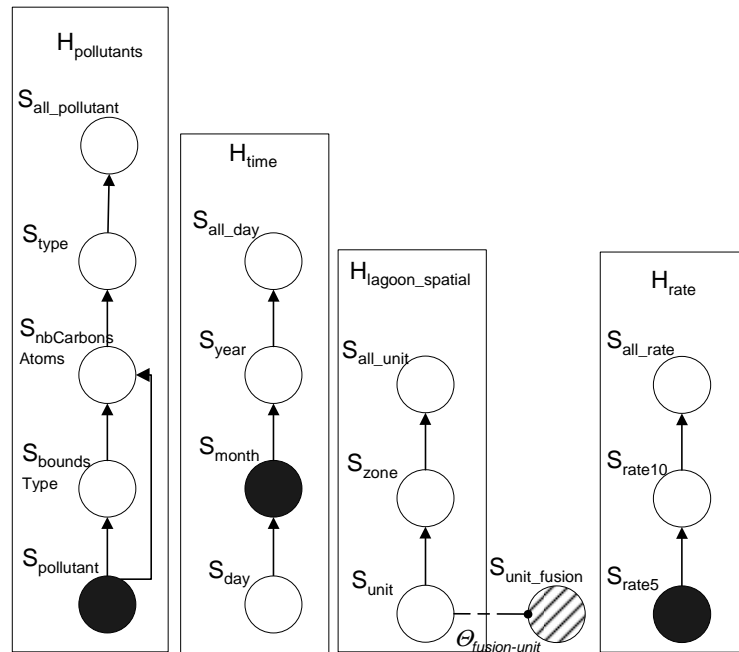


Figure 5.11. Représentation graphique du Schéma de Vue  $V_{\text{corila-unit-month}}$ .

$S_{\text{pollutant}}$ .name_pollutant	$S_{\text{day}}$ .day	$S_{\text{unit\_fusion}}$ .name_unit_fusion	$S_{\text{rate5}}$ .value5
Zinc	9-05	Mazzorbo+Tessera+Murano	45-49
Sulfur Trioxide	9-05	Mazzorbo	30-34
Methane	9-05	Mazzorbo+Tessera+Murano+Coa di Latte	30-34
Zinc	9-05	Canal Fondello	30-34
Benzo(b)fluoronthene	9-05	Murano	45-49
Methane	9-05	Coa di Latte	30-34
Sulfur Trioxide	9-05	Coa di Latte+ Canal fondello	20-24

Tableau 5.3. Instance de  $V_{\text{corila-unit-month}}$ .

Une Vue (Schema et Instance), en utilisant les concepts de Cube de Base et de Mode d'Agrégation, lève la symétrie entre les mesures et les dimensions qui caractérise le Cube de Base, pour définir une requête multidimensionnelle (et son résultat) qui porte sur une mesure géographique ou complexe (agrégée avec le Mode d'Agrégation) et qui utilise un ensemble des niveaux des hiérarchies du Cube de Base.

### 5.3 Algèbre

Dans cette section, nous décrivons l'algèbre associée au modèle de données présenté dans la section 1. Tous les opérateurs de l'algèbre s'appliquent à une Vue et retournent une autre Vue. L'algèbre fournit l'opérateur Roll-Up qui permet de monter sur une dimension, l'opérateur Dice qui élimine une dimension d'analyse, et l'opérateur Slice qui coupe une

partie de l'hypercube. L'algèbre présente aussi deux opérateurs qui changent la granularité de la mesure (Classify et Specialize), l'opérateur Permute qui intervertit mesure et dimension, et l'opérateur OLAP-Buffer qui introduit dans un contexte OLAP l'opérateur d'analyse spatiale de buffer.

### 5.3.1 Opérateurs de forage et de coupe

Dans cette section, nous introduisons les opérateurs Roll-Up, Dice et Slice. L'opérateur de Roll-Up permet d'agréger des objets géographiques ou complexes en montant sur une dimension, contrairement aux approches classiques où les mesures sont de simples valeurs quantitatives. L'opérateur Dice reformule le modèle de l'application en enlevant une dimension. Le Slice permet de couper l'hypercube en utilisant des prédicats alphanumériques et/ou spatiaux définis sur les membres d'une dimension (de manière similaire aux modèles SOLAP) et aussi sur les mesures (détaillées ou agrégées).

#### 5.3.1.1 Roll-Up

L'opérateur *Roll-Up*, noté  $\rho$ , permet de monter dans une hiérarchie et d'agréger les mesures.

Soient  $V_v$  un Schéma de Vue =  $\langle BC_{bc}, \mathcal{L}, \Theta_f, \gamma \rangle$  et  $S_t$  un Schéma d'Entité qui appartient à une hiérarchie du Cube de Base de  $V_v$ . L'opérateur Roll-Up peut être appliqué à  $V_v$  pour « monter » sur  $S_t$ . Le résultat de cette opération est un autre Schéma de Vue  $V'_v = \langle BC_{bc}, \mathcal{L}', \Theta_f, \gamma' \rangle$ . Le Cube de Base et le Mode d'Agrégation ne changent pas et les Niveaux de la Vue sont les mêmes que  $V_v$ , à part pour le nouveau niveau  $S_t$ .

Les instances de  $V'_v$  sont calculées en utilisant l'algorithme Roll-UpAggregation.

L'idée de l'algorithme est de considérer tous les tuples instances possibles  $\langle t_1, \dots, t_m, t_{af} \rangle$  de  $V'_v$ . Si parmi les tuples instances de  $BC_{bc}$  il existe un ensemble de tuples  $\{\langle t_{b1}^j, \dots, t_{b_{m+1}}^j \rangle, \text{ où } j=1 \dots n\}$  où chaque instance  $t_{b_i}^j$ , qui n'est pas une mesure, est un descendant d'une instance  $t_k$ , et si les mesures de ces tuples  $t_{b_1}^1 \dots t_{b_n}^n$  sont telles que  $t_{af}$  est l'agrégation (est agrégat de) de  $t_{b_1}^1 \dots t_{b_n}^n$ , alors la fonction booléenne  $\gamma'$  pour le tuple  $\langle t_1, \dots, t_m, t_{af} \rangle$  est vraie.

*Soient*

- $V_v$  un Schéma de Vue =  $\langle BC_{bc}, \mathcal{L}, \Theta_k, \gamma \rangle$
- $S_t$  un Schéma d'Entité tel que  $S_{L_s} \#_h S_t$  où  $h=H(BC_{bc}, S_t)=H(BC_{bc}, S_{L_s})$

( $S_t$  un Schéma d'Entité qui appartient au même Schéma de Hiérarchie du s-ième Schéma d'Entité de  $\mathcal{L}(S_{L_s})$ )

alors  $\rho(V_v)[S_t] = V'_v = \langle BC_{bc}, \mathcal{L}', \Theta_f, \gamma' \rangle$  où:

- $\mathcal{L}' = \langle S'_{L_1}, \dots, S'_{L_m} \rangle$  où :
  - $\forall i \in [1, \dots, s[ \cup ]s, \dots, m], S'_{L_i} = S_{L_i}$
  - $S'_{L_s} = S_t$
- $\gamma': I(S'_{L_1}) \times \dots \times I(S'_{L_m}) \times I(S_{af})$  où :
  - $S_{af}$  est la mesure agrégée de  $\Theta_k$
  - $\gamma'$  est calculée en utilisant l'algorithme Roll-UpAggregation.

Algorithme Roll-UpAggregation
<p><b>Entrée</b>  <math>V_v</math> : Schéma de Vue sur lequel l'opérateur Roll-Up est appliqué  <math>dom(\gamma') : I(S'_{L1}) \times \dots \times I(S'_{Lm}) \times I(S_{af})</math></p> <p><b>Résultat</b>  <math>\gamma'</math> : fonction booléenne de <math>V'_v</math></p>
<p><i>Soit <math>S_f \in L(H_{bf})</math> //Supposons que le niveau bottom de la hiérarchie <math>H_{bf}</math> du Cube de Base soit utilisé comme mesure</i>  <i>Soit list un ensemble d'instances d'Entité //list va contenir l'ensemble de mesures détaillées associées aux fils du tuple <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> de <math>V'_v</math></i></p> <p><b>for each</b> <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> <b>in</b> <math>dom(\gamma')</math> { // Liste tous les tuples instances possibles de <math>V'_v</math></p> <p style="padding-left: 2em;"><i>list := <math>\emptyset</math></i></p> <p style="padding-left: 2em;"><b>for each</b> <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> <b>in</b> <math>dom(\delta)</math> { //Liste tous les tuples instances de <math>BC_{bc}</math></p> <p style="padding-left: 4em;"><math>\forall i \in [1, \dots, m]</math>, soit <math>t'_{bi} = t_{bj}</math> tel que <math>t_{bj} \in \text{leafs}(H(BC_{bc}, S'_{Li}), \text{all})</math> // <math>t'_{bi}</math> est l'unique instance parmi <math>\{t_{b1}, \dots, t_{bm+1}\}</math> qui appartient à la même hiérarchie que <math>t_i</math></p> <p style="padding-left: 4em;"><b>if</b> <math>((\forall i \in [1, \dots, m], t'_{bi} \in \text{leafs}(H(BC_{bc}, S'_{Li}), t_i)) \text{ and } \delta(t_{b1}, \dots, t_{bm+1}) = \text{true} \text{ and } t_{bf} \notin \text{list})</math></p> <p style="padding-left: 4em;"><b>then</b> add <math>t_{bf}</math> to list // Contrôle que le tuple instance <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> du Cube de Base <math>BC_{bc}</math> est utilisé pour la construction de la mesure agrégée <math>t_{af}</math> appartenant au tuple <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> de <math>V'_v</math></p> <p style="padding-left: 2em;">}</p> <p style="padding-left: 2em;"><b>if</b> <math>(t_{af} \text{ est agrégat de list})</math> //Contrôle que tous les tuples nécessaires pour construire la mesure agrégée sont présents parmi les tuples instances de <math>BC_{bc}</math></p> <p style="padding-left: 2em;"><b>then</b> <math>\gamma'(t_1, \dots, t_m, t_{af}) := \text{true}</math></p> <p style="padding-left: 2em;"><b>else</b> <math>\gamma'(t_1, \dots, t_m, t_{af}) := \text{false}</math></p> <p style="padding-left: 2em;">}</p> <p><b>return</b> <math>(\gamma')</math></p>

#### Exemple 14.

Supposons que l'utilisateur, en utilisant la Vue de Base  $V_{\text{corila-unit}}$ , veuille agréger les unités en montant au niveau mois sur la hiérarchie temporelle. Il utilise l'opérateur Roll-up de la façon suivante :

$\rho(V_{\text{corila-unit}})[S_{\text{month}}] = V_{\text{corila-unit-month}} = \langle BC_{\text{corila}}, \langle S_{\text{pollutant}}, S_{\text{month}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-unit}}, \gamma' \rangle$   
 où  $\gamma' : I(S_{\text{pollutant}}) \times I(S_{\text{month}}) \times I(S_{\text{rate5}}) \times I(S_{\text{unit\_fusion}})$

Cette Vue représente les régions de la lagune polluées pour chaque polluant, pour chaque mois et pour chaque valeur de pollution.

Nous notons que dans  $V_{\text{corila-unit-month}}$ , le Schéma d'Entité  $S_{\text{day}}$  appartenant aux Niveaux de la Vue et à la fonction booléenne de  $V_{\text{corila-unit}}$  est remplacé par  $S_{\text{month}}$ .

L'instance de  $V_{\text{corila-unit-month}}$  est représentée dans le Tableau 5.3. Pour calculer  $\gamma'$  l'algorithme Roll-UpAggregation fonctionne de la façon suivante : il considère tous les tuples instances possibles de  $V_{\text{corila-unit-month}}$ , par exemple  $t_1 = \langle \text{sulfure trioxide}, 9-05, 20-24, \text{Coa di Latte+Canal Fondello+Tessera} \rangle$ . Pour ce tuple, l'algorithme cherche parmi les tuples instances du Cube de Base  $BC_{\text{corila}}$ , les tuples avec le sulfure trioxide, un jour du mois de septembre 2005, une unité et une valeur de pollution de 20-24 mg/l. Les tuples de la table de faits ( $I(BC_{\text{corila}})$ ) qui respectent ces contraintes sont :  $\langle \text{sulfure trioxide}, 26-9-05, \text{Coa di Latte}, 20-24 \rangle$  et  $\langle \text{sulfure trioxide}, 26-9-05, \text{Canal Fondello}, 20-24 \rangle$ . Puisque Coa di Latte+Canal Fondello+Tessera est l'agrégation de Coa di Latte, Fondello et Tessera et non de Coa di Latte et Canal Fondello, la fonction booléenne pour le tuple  $t_1$  est fausse. En revanche, pour le tuple  $t_2 = \langle \text{sulfure trioxide}, 9-05, 20-24, \text{Coa di Latte+Canal Fondello} \rangle$  la fonction booléenne est vraie, car les deux unités Coa di Latte et Canal Fondello, utilisées pour créer la mesure agrégée, sont présentes dans la table des faits.

Il est important de souligner que puisque les hiérarchies peuvent être complexes et une relation plusieurs à plusieurs peut exister entre faits et dimensions, il est possible de prendre en compte plusieurs fois une mesure dans le processus d'agrégation. L'algorithme Roll-UpAggregation, pour garantir une agrégation correcte, utilise toujours les mesures détaillées de la table de faits et contrôle qu'une mesure détaillée soit prise en compte une seule fois à travers la condition  $t_{\text{bf}} \notin \text{list}$ .

### 5.3.1.2 Slice

L'opérateur *Slice*, noté  $\sigma$ , permet de couper l'hypercube en utilisant les membres des dimensions, les mesures détaillées, ou les mesures agrégées.

Soient  $V_v$  un Schéma de Vue =  $\langle BC_{\text{bc}}, \mathcal{L}, \Theta_f, \gamma \rangle$  et  $\pi$  un prédicat. L'opérateur *Slice* peut alors être appliqué à  $V_v$  en utilisant  $\pi$ . Le résultat de cette opération est un autre Schéma de Vue  $V'_v = \langle BC'_{\text{bc}}, \mathcal{L}, \Theta_f, \gamma' \rangle$ . Les schémas hiérarchiques du Cube de Base  $BC'_{\text{bc}}$ , les Schémas d'Entité de  $\mathcal{L}$  et le Mode d'Aggrégation restent les mêmes de  $V_v$ , car le *Slice* n'implique pas de changement de cuboïde dans le treillis des cuboïdes. En revanche, la fonction booléenne de  $BC'_{\text{bc}}$  change, et au cas où  $V_v$  n'est pas une Vue de Base, la fonction booléenne de  $V'_v$  change également.

En particulier, quand le *Slice* est appliqué à une Vue de Base, la fonction booléenne du Cube de Base de  $V_v$  est modifiée de façon à garder seulement les tuples qui satisfont le

prédicat  $\pi$ . Sinon, la fonction booléenne  $\gamma$  du Schéma de Vue  $V_v$  est aussi modifiée de façon à garder seulement les tuples qui satisfont le prédicat  $\pi$ . La fonction booléenne de  $BC_{bc}$  est ensuite modifiée, grâce à l'algorithme SliceBaseCube, pour maintenir une cohérence entre le Cube de Base et la Vue, i.e. l'instance d'une Vue doit toujours être calculée à partir de l'instance du Cube de Base. De cette façon, l'utilisateur peut appliquer une opération de Roll-Up après une opération de Slice et la Vue résultante gardera la coupe effectuée par le Slice. Grâce à cette particularité, l'opérateur de Slice nous permet de définir des requêtes multidimensionnelles sur des membres cousins, comme cela est décrit dans le paragraphe 5.5.2.

L'idée de l'algorithme SliceBaseCube est de considérer tous les tuples instances  $\langle t_1, \dots, t_m, t_{af} \rangle$  de  $V_v$ . Si le tuple  $\langle t_{b1}, \dots, t_{bm+1} \rangle$  où chaque instance  $t_{bi}$ , qui n'est pas une mesure, est un descendant d'une instance  $t_k$ , existe dans la table des faits,  $(I(BC_{bc}))$ , alors la fonction booléenne du nouveau Cube de Base  $BC'_{bc}$  pour  $\langle t_{b1}, \dots, t_{bm+1} \rangle$  est vraie.

*Soient*

- $V_v$  un Schéma de Vue =  $\langle BC_{bc}, \mathcal{L}, \Theta_k, \gamma \rangle$
- $\pi$  un prédicat défini sur  $I(S_t)$  où :
  - $S_t = S_{L_s}$  ou
  - $S_t$  est la mesure détaillée  $S_f$  du Mode d'Agrégation  $\Theta_k$  si  $V_v$  est une Vue de Base ou
  - $S_t$  est la mesure agrégée  $S_{af}$  du Mode d'Agrégation  $\Theta_k$
- si  $V_v$  n'est pas une Vue de Base, alors  $\sigma(V_v)[\pi] = V'_v = \langle BC'_{bc}, \mathcal{L}, \Theta_k, \gamma' \rangle$  où :
  - $\gamma' : I(S_{L1}) \times \dots \times I(S_{Lm}) \times I(S_{af})$  est calculée en utilisant l'algorithme SliceVue.
  - $BC'_{bc} = \langle H_{b1}, \dots, H_{bm+1}, \delta' \rangle$  où  $\delta'$  est définie sur  $I(L_{b1}) \times \dots \times I(L_{bm+1})$  et  $\delta'$  est calculée en utilisant l'algorithme SliceBaseCube
- si  $V_v$  est une Vue de Base, alors  $\sigma(V_v)[\pi] = V'_v = \langle BC'_{bc}, \mathcal{L}, \Theta_j, \gamma \rangle$  où :
  - $BC'_{bc} = \langle H_{b1}, \dots, H_{bm+1}, \delta' \rangle$  où  $\delta'$  est définie sur  $I(L_{b1}) \times \dots \times I(L_{bm+1})$  et  $\delta'$  est calculée en utilisant l'algorithme SliceBasicVue.

<p><b>Algorithme SliceVue</b></p> <p><b>Entrée</b>  <math>V_v</math> : Schéma de Vue auquel l'opérateur Slice est appliqué  <math>\pi</math> : prédicat utilisé pour couper l'hypercube</p> <p><b>Résultat</b>  <math>\gamma</math> : fonction booléenne de <math>V'_v</math></p> <pre> <b>for each</b> <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> <b>in</b> <math>dom(\gamma)</math> { // Liste tous les tuples instances de <math>V_v</math>      Soit <math>t'_t \in I(S_t)</math> tel que <math>t'_t \in \{t_1, \dots, t_m, t_{af}\}</math> // <math>t'_t</math> est un membre d'une dimension     ou <math>t'_t</math> est une mesure agrégée     <b>if</b> (<math>\gamma(t_1, \dots, t_m, t_{af}) = \mathbf{true}</math> and <math>\pi(t'_t) = \mathbf{true}</math>) // Contrôle que le tuple     instance <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> de <math>V_v</math> satisfait le prédicat <math>\pi</math>          <b>then</b> <math>\gamma(t_1, \dots, t_m, t_{af}) := \mathbf{true}</math> <b>else</b> <math>\gamma(t_1, \dots, t_m, t_{af}) := \mathbf{false}</math>     } <b>return</b> (<math>\gamma</math>)                 </pre>
<p><b>Algorithme SliceBaseCube</b></p> <p><b>Entrée</b>  <math>V_v</math> : Schéma de Vue auquel l'opérateur Slice est appliqué  <math>\gamma</math> : fonction booléenne calculée par SliceVue</p> <p><b>Résultat</b>  <math>\delta</math> : fonction booléenne du Cube de Base de <math>V'_v</math></p> <pre> <b>for each</b> <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> <b>in</b> <math>dom(\gamma)</math> { // Liste tous les tuples instances de <math>V'_v</math>      <b>for each</b> <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> <b>in</b> <math>dom(\delta)</math> { // Liste tous les tuples instances     de <math>BC_{bc}</math>          <math>\forall i \in [1, \dots, m]</math>, soit <math>t'_{bi} = t_{bj}</math> tel que <math>t_{bj} \in \text{leafs}(H(BC'_{bc}, S_{Li}), all)</math> // <math>t'_{bi}</math>         est l'unique instance parmi <math>\{t_{b1}, \dots, t_{bm+1}\}</math> qui appartient à la         même hiérarchie que <math>t_i</math>          <b>if</b> (<math>(\forall i \in [1, \dots, m]) t'_{bi} \in \text{leafs}(H(BC'_{bc}, S_{Li}), t_i)</math> and <math>\gamma(t_1, \dots, t_m, t_{af}) =</math>         <b>true</b> and <math>\delta(t_{b1}, \dots, t_{bm+1}) = \mathbf{true}</math>) // Contrôle que le tuple         instance <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> du Cube de Base <math>BC_{bc}</math> est utilisé pour         former la mesure agrégée <math>t_{af}</math> appartenant au tuple instance         <math>\langle t_1, \dots, t_m, t_{af} \rangle</math> de <math>V'_v</math>             <b>then</b> <math>\delta(t_{b1}, \dots, t_{bm+1}) := \mathbf{true}</math>             <b>else</b> <math>\delta(t_{b1}, \dots, t_{bm+1}) := \mathbf{false}</math>         }     } <b>return</b> (<math>\delta</math>)                 </pre>



<b>Algorithme SliceBasicVue</b>
<p><b>Entrée</b>  <math>V_v</math> : Vue de Base à la quelle l'opérateur Slice est appliqué  <math>\pi</math> : prédicat utilisé pour couper l'hypercube</p> <p><b>Résultat</b>  <math>\delta'</math> : fonction booléenne du Cube de Base de <math>V_v</math></p>
<pre> <b>for each</b> <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> <b>in</b> <math>dom(\delta)</math> { // Liste tous les tuples instances du Cube de Base <math>BC_{bc}</math>      Soit <math>t_t \in I(S_t)</math> tel que <math>t_t \in \{t_{b1}, \dots, t_{bm+1}\}</math> // <math>t'_t</math> est un membre d'une dimension ou il est une mesure      <b>if</b> (<math>\delta(t_{b1}, \dots, t_{bm+1}) = \mathbf{true}</math> and <math>\pi(t_t) = \mathbf{true}</math>) // Contrôle que le tuple instance <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> de <math>BC_{bc}</math> satisfait le prédicat      <b>then</b> <math>\delta'(t_{b1}, \dots, t_{bm+1}) := \mathbf{true}</math> <b>else</b> <math>\delta'(t_{b1}, \dots, t_{bm+1}) := \mathbf{false}</math> } <b>return</b> (<math>\delta'</math>) </pre>

**Exemple 15.**

Supposons que l'utilisateur veuille analyser les unités avec une surface supérieure à 1000 m<sup>2</sup>, alors il applique l'opérateur Slice, en utilisant un prédicat spatial sur les mesures de la façon suivante :

$$\sigma(V_{corila-unit})[S_{unit}.f^{area} > 1000] = V_{corila-unit-1000} = \langle BC_{corila1000}, \langle S_{pollutant}, S_{day}, S_{rate5} \rangle, \Theta_{fusion-unit}, \gamma \rangle$$

où  $BC_{corila1000} = \langle H_{pollutants}, H_{time}, H_{lagoon\_spatial}, H_{rate}, \delta' \rangle$  où  $\delta'$  est définie sur  $I(S_{pollutant}) \times I(S_{day}) \times I(S_{unit}) \times I(S_{rate5})$

Les tuples instances de  $V_{corila-unit-1000}$  sont représentés dans le Tableau 5.4.

$S_{pollutant.name\_pollutant}$	$S_{day.day}$	$S_{unit.name\_unit}$	$S_{rate5.value5}$
Zinc	19-9-05	Mazzorbo	45-49
Sulfur Trioxide	19-9-05	Mazzorbo	30-34
Methane	19-9-05	Mazzorbo	30-34
Zinc	19-9-05	Canal Fondello	30-34
Benzo(b)fluoronthene	19-9-05	Murano	45-49
Methane	19-9-05	Murano	30-34
Sulfur Trioxide	26-9-05	Canal Fondello	20-24
Zinc	19-9-05	Murano	45-49

Tableau 5.4. Instance de  $V_{corila-unit-1000}$ .

**Exemple 16.**

Supposons que l'utilisateur veuille analyser les zones polluées et avec une surface supérieure à 1500 m<sup>2</sup> pour chaque mois. Il applique alors l'opérateur de Slice en utilisant un prédicat spatial sur les mesures agrégées :

$$\sigma(V_{\text{corila-unit-month}})[S_{\text{unit\_fusion}}.f^{\text{area}} > 1500] = V_{\text{corila-unit-month1500}} = \langle BC_{\text{corila1500}}, \langle S_{\text{pollutant}}, S_{\text{month}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-unit}}, \gamma \rangle$$

où  $BC_{\text{corila1500}} = \langle H_{\text{pollutants}}, H_{\text{time}}, H_{\text{lagoon\_spatial}}, H_{\text{rate}}, \delta' \rangle$  où  $\delta'$  est définie sur  $I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S_{\text{unit}}) \times I(S_{\text{rate5}})$

Les tuples instances de  $V_{\text{corila-unit-month-1500}}$  sont représentés dans le Tableau 5.5.

L'instance de  $BC_{\text{corila1500}}$  est représentée dans le Tableau 5.6.

Pour calculer l'instance de  $BC_{\text{corila1500}}$ , l'algorithme SliceBaseCube fonctionne de la façon suivante : il considère toutes les tuples instances de  $V_{\text{corila-unit-month-1500}}$ , par exemple  $t_1 = \langle \text{sulfure trioxide}, 9-05, 20-24, \text{Coa di Latte+Canal Fondello} \rangle$ . Pour ce tuple, l'algorithme cherche parmi les tuples instances du Cube de Base  $BC_{\text{corila}}$ , les tuples avec le sulfure trioxide, un jour du mois de septembre 2005, une unité et une valeur de pollution égal à 20-24 mg/l. Les tuples présents dans la table de faits qui respectent ces contraintes sont  $\langle \text{sulfure trioxide}, 26-9-05, \text{Coa di Latte}, 20-24 \rangle$  et  $\langle \text{sulfure trioxide}, 26-9-05, \text{Canal Fondello}, 20-24 \rangle$ . Alors, pour ces deux tuples la fonction booléenne du Cube de Base  $BC_{\text{corila1500}}$  est vraie.

Nous notons que les tuples instances du Cube de Base  $BC_{\text{corila1500}}$  et de la Vue  $V_{\text{corila-unit-month-1500}}$  restent cohérents après l'opération de Slice, ou en d'autres termes qu'une opération de Roll-Up sur les mois à partir de la Vue de Base qu'utilise  $BC_{\text{corila1500}}$  créera une Vue dont les instances sont celles de  $V_{\text{corila-unit-month-1500}}$ .

$S_{\text{pollutant}}.name\_pollutant$	$S_{\text{month}}.month$	$S_{\text{unit\_fusion}}.name\_unit\_fusion$	$S_{\text{rates}}.value5$
Zinc	9-05	Mazzorbo+Tessera+Murano	45-49
Methane	9-05	Mazzorbo+Tessera+Murano+Coa di latte	30-34
Sulfur Trioxide	9-05	Coa di Latte+ Canal fondello	20-24

Tableau 5.5. Instance de  $V_{\text{corila-unit-month-1500}}$ .

$S_{\text{pollutant.name\_pollutant}}$	$S_{\text{day.day}}$	$S_{\text{unit.name\_unit}}$	$S_{\text{rate5.value5}}$
Zinc	19-9-05	Mazzorbo	45-49
Methane	19-9-05	Mazzorbo	30-34
Methane	19-9-05	Murano	30-34
Zinc	19-9-05	Tessera	45-49
Methane	19-9-05	Coa di Latte	30 -34
Zinc	19-9-05	Murano	45-49
Sulfur Trioxide	26-9-05	Coa di Latte	20-24
Methane	26-9-05	Tessera	30-34
Sulfur Trioxide	26-9-05	Canal Fondello	20-24

Tableau 5.6. Instance de  $BC_{\text{corila1500}}$ .

L'opérateur de Slice permet ainsi de couper l'hypercube en utilisant les membres des dimensions, les mesures ou les mesures agrégées.

### 5.3.1.3 Dice

L'opérateur Dice, noté  $\eta$ , supprime un axe d'analyse.

Soient  $V_v$  une Vue de Base =  $\langle BC_{bc}, \mathcal{L}, \Theta_f, \gamma \rangle$  et  $H_{bt}$  un Schéma de Hiérarchie du Cube de Base de  $V_v$ , alors l'opérateur Dice peut être appliqué à  $V_v$  en utilisant  $H_{bt}$ . Le résultat de cette opération est un autre Schéma de Vue  $V'_v = \langle BC'_{bc}, \mathcal{L}', \Theta_f, \gamma' \rangle$ . Le Mode d'Agrégation ne change pas. Le Cube de Base  $BC'_{bc}$  ne présente pas la hiérarchie  $H_{bt}$  et les Niveaux de la Vue ne contiennent plus le Schéma d'Entité appartenant à  $H_{bt}$ .

La fonction booléenne du nouveau Cube de Base  $BC'_{bc}$  de  $V'_v$  est calculée en utilisant l'algorithme CutHierarchy.

L'algorithme considère tous les tuples instances  $\langle t_1, \dots, t_m, t_{af} \rangle$  de  $V_v$  et assigne vrai à la fonction booléenne du nouveau Cube de Base de  $V'_v$  pour le tuple  $\langle t_1, \dots, t_{t-1}, t_{t+1}, \dots, t_m, t_{af} \rangle$ .

Soient :

- $V_v$  une Vue de Base =  $\langle BC_{bc}, L, \Theta_k, \gamma \rangle$
- $H_{bt}$  un Schéma de Hiérarchie de  $BC_{bc}$

alors  $\eta(V_v) [H_{bt}] = V'_v = \langle BC'_{bc}, L', \Theta_f, \gamma' \rangle$  où :

- $L' = \langle S'_{L1}, \dots, S'_{Lm-1} \rangle$  où :
- Si  $S_{Lj} \in L(H_{bt})$ , alors  $\forall i \in [1, \dots, j-1], S'_{Li} = S_{Li}$  et  $\forall i \in [j, \dots, m-1], S'_{Li} = S_{Li+1}$
- $\gamma' : I(S'_{L1}) \times \dots \times I(S'_{Lm}) \times I(S_{af})$
- $BC'_{bc} = \langle H_{bt1}, \dots, H_{btm}, \delta' \rangle$  où :
- $\forall i \in [1, \dots, t-1], H_{bt'i} = H_{bt_i}$  et  $\forall i \in [t, \dots, m], H_{bt'i} = H_{bt_{i+1}}$
- $\delta' : I(L_{bt'1}) \times \dots \times I(L_{bt'm})$  est calculée en utilisant l'algorithme CutHierarchy

Algorithme CutHierarchy
<p><b>Entrée</b>  <math>V_v</math> : Vue de Base à laquelle l'opérateur Dice est appliqué  <math>H_{bt}</math> : hiérarchie qui doit être éliminée</p> <p><b>Résultat</b>  <math>\delta'</math> : fonction booléenne du Cube de Base de <math>V'_v</math></p>
<pre> <b>for each</b> <math>\langle t_{bt1}, \dots, t_{btm+1} \rangle</math> <b>in</b> <math>dom(\delta)</math> { // Liste tous les tuples instances du Cube de Base <math>BC_{bc}</math>    <math>\forall i \in [1, \dots, t-1]</math> soit <math>t'_{bt_i} = t_{bt_i}</math>   <math>\forall i \in [t, \dots, m]</math> soit <math>t'_{bt_i} = t_{bt_{i+1}}</math>    <b>if</b> <math>\delta(t_{bt1}, \dots, t_{btm+1}) = \mathbf{true}</math>   <b>then</b> <math>\delta'(t'_{bt1}, \dots, t'_{btm}) := \mathbf{true}</math>   <b>else</b> <math>\delta'(t'_{bt1}, \dots, t'_{btm}) := \mathbf{false}</math> } <b>return</b> (<math>\delta'</math>) </pre>

### Exemple 17.

Supposons que le décideur désire analyser les unités polluées en fonction des polluants et des valeurs de pollution seulement. Alors, il applique l'opération de Dice sur la hiérarchie représentant le temps de la façon suivante :

$\eta(V_{\text{corila-unit}}) [H_{\text{time}}] = V_{\text{corila-unit/Time}} = \langle BC_{\text{corila/Time}}, \langle S_{\text{pollutant}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-unit}}, \gamma' \rangle$  où :

$BC_{\text{corila/Time}} = \langle H_{\text{pollutants}}, H_{\text{lagoon\_spatial}}, H_{\text{rate}}, \delta' \rangle$  où  $\delta' : I(S_{\text{pollutant}}) \times I(S_{\text{unit}}) \times I(S_{\text{rate5}})$

Le nouveau Schéma de Vue est représenté en Figure 5.12. L'instance de  $V_{\text{corila-unit/Time}}$  est représentée dans le Tableau 5.7.

Pour calculer l'instance de  $V_{\text{corila-unit/Time}}$  l'algorithme CutHierarchy fonctionne de la manière suivante : il considère tous les tuples instances de  $V_{\text{corila-unit}}$ , par exemple  $t_1 = \langle \text{sulfure trioxide}, 26-9-05, \text{Coa di Latte}, 20-24 \rangle$ . Alors, la fonction booléenne pour le tuple  $\langle \text{sulfure trioxide}, \text{Coa di Latte}, 20-24 \rangle$  de  $V_{\text{corila-unit/Time}}$  est vraie, car ce tuple contient les mêmes instances que  $t_1$  sauf l'instance représentant le jour 26-9-05.

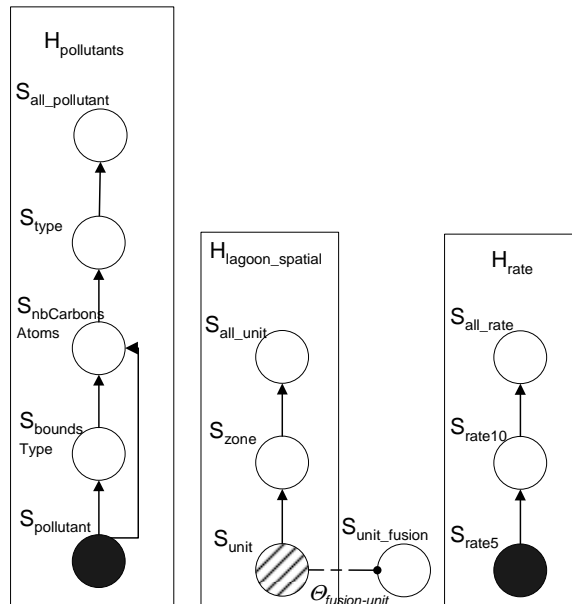


Figure 5.12. Représentation graphique de  $V_{\text{corila-unit/Time}}$ .

$S_{pollutant.name\_pollutant}$	$S_{unit.name\_unit}$	$S_{rate5.value5}$
Zinc	<i>Mazzorbo</i>	45-49
Sulfur Trioxide	<i>Mazzorbo</i>	30-34
Methane	<i>Mazzorbo</i>	30-34
Zinc	<i>Canal Fondello</i>	30-34
Benzo(b)fluoronthene	<i>Murano</i>	45-49
Methane	<i>Murano</i>	30-34
Zinc	<i>Tessera</i>	45-49
Methane	<i>Coa di Latte</i>	30 -34
Zinc	<i>Murano</i>	45-49
Sulfur Trioxide	<i>Coa di Latte</i>	20-24
Methane	<i>Tessera</i>	30-34
Sulfur Trioxide	<i>Canal Fondello</i>	20-24

Tableau 5.7. Instance de  $V_{corila-unit/Time}$ .

### 5.3.2 Opérateurs de navigation dans la hiérarchie de la mesure

Dans notre approche, une mesure peut appartenir à une hiérarchie, qui représente les relations spatiales, les relations de généralisation ou les relations non spatiales des objets entre eux. Nous introduisons deux nouveaux opérateurs qui, en utilisant ces hiérarchies, permettent de changer la granularité de la mesure. Ces opérateurs permettent de monter dans la hiérarchie de la mesure et de remplacer un ensemble de mesures détaillées avec des mesures appartenant à un niveau moins détaillé. Les deux opérateurs que nous introduisons sont le Classify et le Specialize. Ils utilisent deux critères différents pour effectuer ce remplacement. Le Classify nécessite que tous les descendants des nouvelles mesures soient présents dans la table de faits pour effectuer le remplacement. Pour le Specialize, un seul descendant est suffisant. Grâce à ces deux politiques différentes de remplacement, les deux opérateurs nous permettent de répondre à des requêtes différentes. Supposons par exemple que l'utilisateur veuille analyser les zones de la lagune telles que toutes leurs unités sont polluées. Il applique alors l'opérateur de Classify, dans le cas où il est intéressé par les zones de la lagune telles qu'au moins une de leurs unités est polluée, alors il utilise le Specialize.

Avant de décrire en détail ces opérateurs, nous présentons le concept de Hiérarchie Réduite. Une mesure détaillée est toujours prise au niveau le plus détaillé d'une hiérarchie dans le Cube de Base. Lors du changement de mesure, la définition du Schéma de Cube de Base de la Vue doit rester cohérente, c'est à dire, la nouvelle mesure doit toujours être le niveau le plus détaillé d'une hiérarchie. L'idée est alors de remplacer la hiérarchie qui contient l'ancienne mesure par une nouvelle hiérarchie qui a comme niveau bottom la nouvelle mesure. Cette hiérarchie est la Hiérarchie Réduite.

Soient  $H_h$  un Schéma de Hiérarchie,  $\uparrow_h$  l'instance de  $H_h$ , et  $S_t$  un Schéma d'Entité qui appartient aux Schémas d'Entité Niveaux de  $H_h$ , alors la Hiérarchie Réduite (Schéma et Instance) est une hiérarchie tel que les niveaux appartiennent au sous-arbre de  $\#_h$  qui a comme feuille  $S_t$ , et les membres appartiennent au sous-arbre de  $\uparrow_h$  qui a comme feuilles les membres du niveau  $S_t$ . Pour calculer la Hiérarchie Réduite nous proposons l'algorithme ReduceHierarchy

Algorithme ReduceHierarchy
<p><b>Entrée</b>  <math>H_h</math> et <math>\uparrow_h</math> : schéma et l'instance de la hiérarchie qui doit être réduite  <math>S_t</math> : niveau de <math>H_h</math> qui doit être utilisé pour transformer la hiérarchie</p> <p><b>Résultat</b>  <math>H'_h</math> et <math>\uparrow_{h'}</math> : schéma et instance de la nouvelle hiérarchie</p>
<p><i>Soit list un ensemble de Schémas d'Entité //list va contenir l'ensemble des niveaux de la nouvelle hiérarchie</i></p> <p><i>list := <math>\emptyset</math></i></p> <p><i>Calculer le sous-arbre de l'arbre représenté par <math>\uparrow_h</math> ayant comme feuilles les instances de <math>S_t</math></i></p> <p><b>for each</b> Schéma d'Entité <math>S_i</math> tel qu'au moins une de ses instances appartient au sous-arbre {              <math>S'_i := S_i</math>              <math>I(S'_i) := I''(S_i)</math> où <math>I''(S_i)</math> est l'ensemble des instances de <math>S_i</math> inclus dans le sous-arbre              <b>if</b> (<math>S'_i = S_t</math>) <b>then</b> <math>\downarrow_{h'} := S'_i</math>              <b>if</b> (<math>S'_i = \sqrt{h}</math>) <b>then</b> <math>\sqrt{h'} := S'_i</math>              <b>if</b> (<math>S'_i \neq \sqrt{h'}</math>) <b>and</b> <math>S'_i \neq \downarrow_{h'}</math>) <b>then</b> add <math>S'_i</math> to list          }</p> <p><math>\#_{h'}</math> est défini par : <b>if</b> <math>S_i \#_h S_j</math> <b>then</b> <math>S'_i \#_{h'} S'_j</math></p> <p><math>\uparrow_{h'}</math> est défini par : <b>if</b> (<math>t_i \uparrow_h t_j</math> <b>and</b> <math>t_i \in I(S'_i)</math> <b>and</b> <math>t_j \in I(S'_j)</math>) <b>then</b> <math>t_i \uparrow_{h'} t_j</math></p> <p><math>L_{h'} := list</math></p> <p><b>return</b> (<math>H'_h = \langle L_{h'}, \downarrow_{h'}, \sqrt{h'}, \#_{h'} \rangle, \uparrow_{h'}</math>)</p>

**Exemple 18.**

Le schéma et l'instance de la Hiérarchie Réduite de  $H_a$  (Figure 5.13a) sur  $S_b$  sont présentés en Figure 5.13b.

Nous remarquons que le niveau bottom de la nouvelle hiérarchie est  $S'_b$  qui est égal à  $S_b$ , et que dans cette nouvelle hiérarchie il n'y a plus le niveau  $S_d$  car aucune de ses instances n'est liée aux instances de  $S_b$  dans la hiérarchie originale.

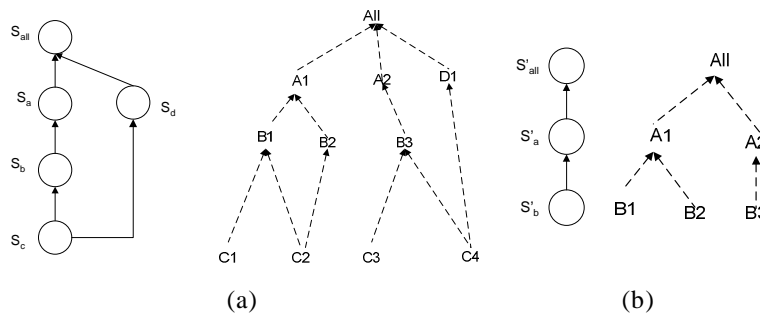


Figure 5.13. a) Schéma de Hiérarchie  $H_a$  et Instance de  $H_a$  b) Hiérarchie Réduite (Schéma et Instance) de  $H_a$  sur  $S_b$ .

**Exemple 19.**

Un exemple de Hiérarchie Réduite est montré en Figure 5.14. Le schéma et l'instance de la Hiérarchie Réduite de  $H_{lagoon\_spatial}$  sur  $S_{zone}$  sont montrés respectivement en Figure 5.14a et 5.14b.

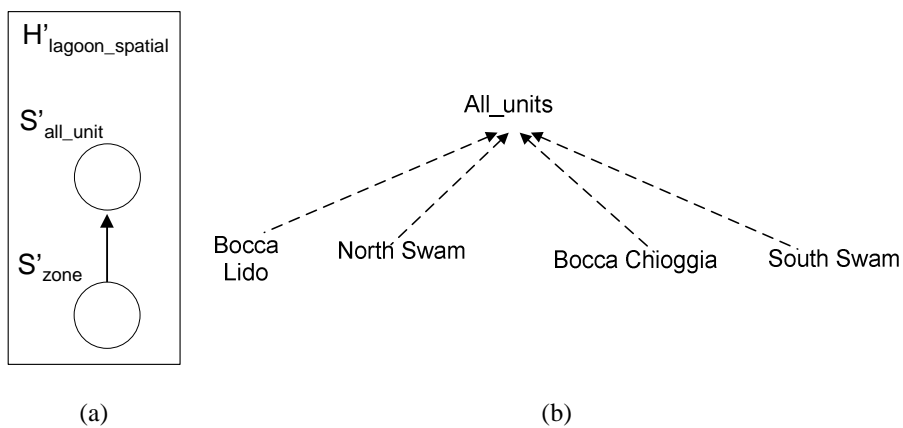


Figure 5.14. Hiérarchie Réduite de  $H_{lagoon\_spatial}$  sur  $S_{zone}$  a) Schéma b) Instance.

**5.3.2.1 Classify**

L'opérateur *Classify*, noté  $\omega$ , permet de changer la granularité de la mesure. Il transforme une Vue dans une nouvelle Vue avec une nouvelle mesure.



Soient  $V_v$  un Schéma de Vue =  $\langle BC_{bc}, \mathcal{L}, \Theta_t, \gamma \rangle$  et  $\Theta_t = \langle S_t, S_{at}, \Phi \rangle$  un Mode d'Agrégation tel que  $S_t$  appartient à la même hiérarchie de la mesure de  $V_v$ , Classify est appliqué à  $V_v$  en utilisant  $\Theta_t$ . Le résultat est un autre Schéma de Vue  $V'_v = \langle BC'_{bc}, \mathcal{L}, \Theta_t, \gamma' \rangle$ . Contrairement au Roll-Up, les niveaux de  $V'_v$  ne changent pas. Ce qui change ce sont toutes les informations liées à la mesure : le Mode d'Agrégation, la fonction booléenne de la Vue, et le Cube de Base. Ce dernier présente une Hiérarchie Réduite et une nouvelle fonction booléenne qui est calculée grâce à l'algorithme Classification.

L'idée de l'algorithme est de considérer tous les tuples instances possibles  $\langle t_1, \dots, t_{m+1} \rangle$  de  $BC'_{bc}$ . Si parmi les tuples instances de  $BC_{bc}$ , existe un ensemble de tuples  $\{\langle t_{b1}^j, \dots, t_{bm+1}^j \rangle, \text{ où } j=1 \dots n\}$  où chaque instance  $t_{bi}^j$ , qui n'est pas une mesure, est égal à  $t_i$ , et si les mesures de ces tuples  $t_{bj}^1, \dots, t_{bj}^n$  sont tous les descendants de  $t_j$  (la nouvelle mesure), alors la fonction booléenne du Cube de Base  $BC'_{bc}$  pour le tuple  $\langle t_1, \dots, t_{m+1} \rangle$  est vraie.

Soient :

- $V_v$  une Vue de Base =  $\langle BC_{bc}, \mathcal{L}, \Theta_k, \gamma \rangle$
- $S_t$  un Schéma d'Entité tel  $S_f \#_h S_t$  où  $h = H(BC_{bc}, S_f) = H(BC_{bc}, S_t)$  et  $S_f$  est la mesure détaillée du Mode d'Agrégation  $\Theta_k$  de  $V_v$  ( $S_t$  appartient à la même hiérarchie que  $S_f$  (la mesure de  $V_v$ ) et  $S_t$  représente un niveau moins détaillé que  $S_f$ )
- Un Mode d'Agrégation  $\Theta_t = \langle S'_t, S_{at}, \Phi \rangle$  où  $S'_t = \downarrow_{h'}$  où  $h'$  est la Hiérarchie Réduite de  $H(BC_{bc}, S_t)$  sur  $S_t$

alors  $\alpha(V_v)[S_b, \Theta_t] = V'_v = \langle BC'_{bc}, \mathcal{L}, \Theta_t, \gamma' \rangle$  où :

- $BC'_{bc} = \langle H_{b1}, \dots, H_{bm+1}, \delta \rangle$  où :
- Si  $S_f \in L(H_{bt})$ , alors  $\forall i \in [1 \dots t[ \cup ]t \dots m+1]$ ,  $H_{bi'} = H_{bi}$  et  $H_{bi'}$  est la Hiérarchie Réduite de  $H(BC_{bc}, S_t)$  sur  $S_t$
- $\delta : I(L_{b1'}) \times \dots \times I(L_{bm+1'})$  est calculée en utilisant l'algorithme Classification
- $\gamma' : I(S_{L1}) \times \dots \times I(S_{Lm}) \times I(S_{at})$ .

<p><b>Algorithmme Classification</b></p> <p><b>Entrée</b>  <math>V_v</math> : Vue de Base à laquelle l'opérateur Classify est appliqué  <math>S_t</math> : niveau à utiliser pour monter dans la hiérarchie de la mesure</p> <p><b>Résultat</b>  <math>\delta</math> : fonction booléenne du Cube de Base de <math>V'_v</math></p> <p><i>Soit <math>S_f \in L(H_{bt})</math> //Le niveau bottom de la hiérarchie <math>H_{bt}</math> du Cube de Base est utilisée comme mesure</i></p> <p><i>Soit list un ensemble d'instances d'Entité //list va contenir l'ensemble de mesures détaillées qui sont les descendants de la nouvelle mesure dans le tuple <math>\langle t_1, \dots, t_{m+1} \rangle</math> de <math>V'_v</math></i></p> <p><b>for each</b> <math>\langle t_1, \dots, t_{m+1} \rangle</math> <b>in</b> <math>dom(\delta)</math> { // Liste tous les tuples instances possibles de <math>BC'_{bc}</math></p> <p style="padding-left: 2em;"><i>list := <math>\emptyset</math></i></p> <p><b>for each</b> <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> <b>in</b> <math>dom(\delta)</math> { //Liste tous les tuples instances de <math>BC_{bc}</math></p> <p style="padding-left: 2em;"><b>if</b> (<math>(\forall i \in [1, \dots, t-1] \cup [t+1, \dots, m+1] t_{bi} = t_i)</math> <b>and</b> <math>\delta(t_{b1}, \dots, t_{bm+1}) = \mathbf{true}</math> <b>and</b> <math>t_{bt} \in \text{leafs}(H(BC_{bc}, S_t), t_t)</math> <b>and</b> <math>t_{bt} \notin \text{list}</math>)</p> <p style="padding-left: 2em;"><b>then</b> add <math>t_{bt}</math> to list //Contrôle que le tuple instance <math>\langle t_1, \dots, t_{m+1} \rangle</math> de <math>BC'_{bc}</math> présente les mêmes instances du tuple instance <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> de <math>BC_{bc}</math> sauf pour la mesure <math>t_{bt}</math> qui doit être un descendant de la nouvelle mesure <math>t_t</math></p> <p style="padding-left: 2em;">}</p> <p style="padding-left: 2em;"><b>if</b> (<math>\text{list} = \text{leafs}(H(BC_{bc}, S_t), t_t)</math> //contrôle que tous les descendants de la nouvelle mesure sont présents dans le Cube de Base <math>BC_{bc}</math></p> <p style="padding-left: 2em;"><b>then</b> <math>\delta(t_1, \dots, t_{m+1}) := \mathbf{true}</math> <b>else</b> <math>\delta(t_1, \dots, t_{m+1}) := \mathbf{false}</math></p> <p style="padding-left: 2em;">}</p> <p><b>return</b> (<math>\delta</math>)</p>
--

**Exemple 20.**

Supposons que le décideur souhaite changer la granularité de la mesure géographique représentée par les unités avec les zones de la lagune, afin de répondre à la requête : « *Quelles sont les zones, dont toutes les unités sont polluées, pour chaque jour et pour chaque polluant ?* ».

Il utilise ainsi l'opérateur Classify :

Soit  $\Theta_{\text{fusion-zone}} = \langle S'_{\text{zone}}, S'_{\text{zone\_fusion}}, \Phi \rangle$  alors  $\omega(V_{\text{corila-unit}})[S_{\text{zone}}, \Theta_{\text{fusion-zone}}] =$

$V_{\text{corila-zone}} = \langle BC_{\text{corila-zone}}, \langle S_{\text{pollutant}}, S_{\text{day}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-zone}}, \gamma' \rangle$  où :

- $BC_{\text{corila-zone}} = \langle H_{\text{pollutants}}, H_{\text{time}}, H'_{\text{lagoon\_spatial}}, H_{\text{rate}} \rangle, \delta'$  où  $\delta'$  est définie sur  $I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S'_{\text{zone}}) \times I(S_{\text{rate5}})$  et  $H'_{\text{lagoon\_spatial}}$  est la Hiérarchie Réduite de  $H_{\text{lagoon\_spatial}}$  sur  $S_{\text{zone}}$
- $\gamma' : I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S_{\text{rate5}}) \times I(S'_{\text{zone\_fusion}})$

La représentation graphique de  $V_{\text{corila-zone}}$  et son instance sont montrées en Figure 5.15 et Tableau 5.8 respectivement. Pour calculer l'instance de  $V_{\text{corila-zone}}$  l'algorithme Classification fonctionne de la façon suivante : il assigne vrai au tuple  $\langle \text{Sulfur Trioxide}, 26-9-05, \text{Bocca Lido}, 20-24 \rangle$ , car les tuples  $\langle \text{Sulfur Trioxide}, 26-9-05, \text{Coa di Latte}, 20-24 \rangle$  et  $\langle \text{Sulfur Trioxide}, 26-9-05, \text{Canal Fondello}, 20-24 \rangle$  sont présents dans la table des faits. En d'autres termes, le 26-9-05 Bocca Lido est polluée par le sulfure trioxide avec une valeur de pollution égal à 20-24 mg/l car ses unités, Coa di Latte et Canal Fondello, sont polluées.

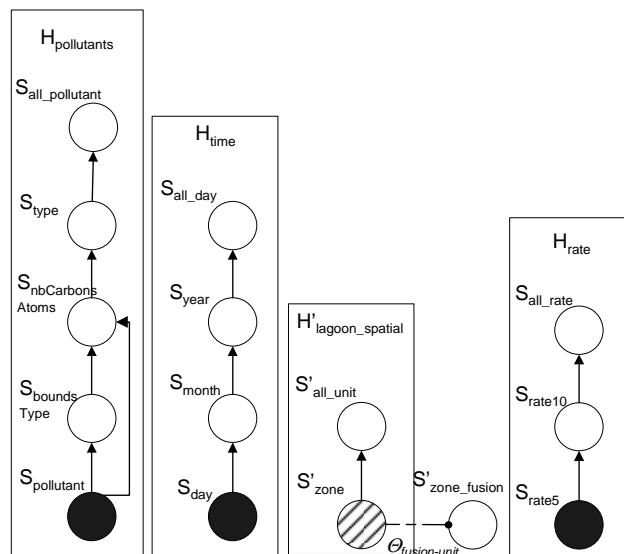


Figure 5.15. Représentation graphique de  $V_{\text{corila-zone}}$ .

$S_{\text{pollutant.name\_pollutant}}$	$S_{\text{day.day}}$	$S'_{\text{zone.name\_zone}}$	$S_{\text{rate5.values5}}$
Zinc	19-9-05	North Swam	45-49
Methane	19-9-05	North Swam	30-34
Sulfur Trioxide	26-9-05	Bocca Lido	20-24

Tableau 5.8. Instance de  $V_{\text{corila-zone}}$ .

**Exemple 21.**

Supposons que l'utilisateur veuille répondre à la requête : « *Pour chaque jour et pour chaque polluant, quelles sont les zones dont toutes les unités ont une surfaces supérieure à 1000 et sont polluées ?* ». Il applique ainsi l'opérateur Classify à la Vue résultante de l'opération de Slice sur les mesures :

Soit  $\Theta_{\text{fusion-zone}} = \langle S'_{\text{zone}}, S'_{\text{zone\_fusion}}, \Phi \rangle$ , alors  $\omega(V_{\text{corila-unit-1000}})[S_{\text{zone}}, \Theta_{\text{fusion-zone}}] = V_{\text{corila-zone-1000}} = \langle BC_{\text{corila-zone-100}}, \langle S_{\text{pollutant}}, S_{\text{day}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-zone}}, \gamma' \rangle$  où:

- $BC_{\text{corila-zone-1000}} \langle H_{\text{pollutants}}, H_{\text{time}}, H'_{\text{lagoon\_spatial}}, H_{\text{rate}} \rangle, \delta' \rangle$  où :
- $\delta': I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S'_{\text{zone}}) \times I(S_{\text{rate5}})$
- $H'_{\text{lagoon\_spatial}}$  est la Hiérarchie Réduite de  $H_{\text{lagoon\_spatial}}$  sur  $S_{\text{zone}}$
- $\gamma': I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S_{\text{rate5}}) \times I(S'_{\text{zone\_fusion}})$

Les tuples instances de  $V_{\text{corila-zone-1000}}$  sont montrés dans le Tableau 5.9.

$S_{\text{pollutant.name\_pollutant}}$	$S_{\text{day.day}}$	$S'_{\text{zone.name\_zone}}$	$S_{\text{rate5.value5}}$
Zinc	19-9-05	North Swam	45-49
Methane	19-9-05	North Swam	30-34

Tableau 5.9. Instance de  $V_{\text{corila-zone-1000}}$ .

Les exemples 20 et 21 montrent une caractéristique importante des opérateurs qui changent la mesure. Les résultats de l'opérateur Classify représentés dans le Tableaux 5.8 et 5.9 sont différents, même si le processus de navigation est le même, car les données utilisées sont différentes (Tableau 5.2 et Tableau 5.4) à cause d'une opération de coupe sur les mesures. Cela signifie que, si une opération de coupe ne peut pas utiliser les mesures et leurs relations hiérarchiques, alors le changement de granularité de la mesure ne peut pas être modélisé à travers une simple navigation parmi plusieurs hypercubes définis a priori, où un hypercube est défini pour chaque niveau de la hiérarchie de la mesure. Dans ce cas, l'hypercube avec la nouvelle mesure doit être recalculé à chaque fois, en utilisant les données de l'hypercube auquel une opération de coupe a été appliquée.

**5.3.2.2 Specialize**

L'opérateur Specialize, noté  $\Lambda$ , permet de changer la granularité de la mesure, en utilisant la même logique que l'opérateur de Classify, avec la seule différence qu'il suffit qu'un seul descendant de la mesure soit présent dans la table des faits, pour effectuer le remplacement. Le Specialize est défini formellement de la façon suivante :

Soient :

- $V_v$  une Vue de Base =  $\langle BC_{bc}, \mathcal{L}, \Theta_k, \gamma \rangle$
- $S_t$  un Schéma d'Entité tel  $S_f \#_h S_t$  où  $h = H(BC_{bc}, S_f) = H(BC_{bc}, S_t)$  et  $S_f$  est la mesure détaillée du Mode d'Agrégation  $\Theta_k$  de  $V_v$  ( $S_t$  appartient à la même hiérarchie que  $S_f$  (la mesure de  $V_v$ ) et  $S_t$  représente un niveau moins détaillé que  $S_f$ )
- Un Mode d'Agrégation  $\Theta_t = \langle S'_t, S_{at}, \Phi \rangle$  où  $S'_t = \mathcal{L}_h$  où  $h$  est la Hiérarchie Réduite de  $H(BC_{bc}, S_t)$  sur  $S_t$

alors  $\mathcal{A}(V_v)[S_t, \Theta_t] = V'_v = \langle BC'_{bc}, \mathcal{L}, \Theta_t, \gamma' \rangle$  où :

- $BC'_{bc} = \langle H_{b1'}, \dots, H_{bm+1'} \rangle$  où :
- Si  $S_f \in \mathcal{L}(H_{bt})$ , alors  $\forall i \in [1 \dots t] \cup ]t \dots m+1]$ ,  $H_{bi'} = H_{bi}$  et  $H_{b't}$  est la Hiérarchie Réduite de  $H(BC_{bc}, S_t)$  sur  $S_t$
- $\delta' : I(\mathcal{L}_{b1'}) \times \dots \times I(\mathcal{L}_{bm+1'})$  est calculée en utilisant l'algorithme Specialization
- $\gamma' : I(S_{L1}) \times \dots \times I(S_{Lm}) \times I(S_{at})$ .

Algorithme Specialization
<p><b>Entrée</b></p> <p><math>V_v</math> : Vue de Base à laquelle l'opérateur Classify est appliqué  <math>S_t</math> : niveau à utiliser pour monter dans la hiérarchie de la mesure</p> <p><b>Résultat</b></p> <p><math>\delta'</math> : fonction booléenne du Cube de Base de <math>V'_v</math></p>
<p>Soit <math>S_f \in L(H_{bt})</math> //Le niveau bottom de hiérarchie <math>H_{bt}</math> du Cube de Base est utilisée comme mesure</p> <p>Soit <math>list</math> un ensemble d'instances d'Entité //list va contenir l'ensemble de mesures détaillées qui sont les descendants de la nouvelle mesure dans le tuple <math>\langle t_1, \dots, t_{m+1} \rangle</math> de <math>V'_v</math></p> <p><b>for each</b> <math>\langle t_1, \dots, t_{m+1} \rangle</math> in <math>dom(\delta')</math> { // Liste tous les tuples instances possibles de <math>BC'_{bc}</math></p> <p style="padding-left: 2em;"><math>list := \emptyset</math></p> <p style="padding-left: 2em;"><b>for each</b> <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> in <math>dom(\delta)</math> { // Liste tous les tuples instances de <math>BC_{bc}</math></p> <p style="padding-left: 4em;"><b>if</b> <math>((\forall i \in [1, \dots, t-1] \cup [t+1, \dots, m+1] t_{bi} = t_i)</math> <b>and</b> <math>\delta(t_{b1}, \dots, t_{bm+1}) = \mathbf{true}</math> <b>and</b> <math>t_{bt} \in \text{leafs}(H(BC_{bc}, S_t), t_t)</math> <b>and</b> <math>t_{bt} \notin list</math>)</p> <p style="padding-left: 4em;"><b>then</b> add <math>t_{bt}</math> to <math>list</math> //Contrôle que le tuple instance <math>\langle t_1, \dots, t_{m+1} \rangle</math> de <math>BC'_{bc}</math> présente les mêmes instances du tuple instance <math>\langle t_{b1}, \dots, t_{bm+1} \rangle</math> de <math>BC_{bc}</math> sauf pour la mesure <math>t_{bt}</math> qui doit être une descendante de la nouvelle mesure <math>t_t</math></p> <p style="padding-left: 2em;">}</p> <p style="padding-left: 2em;"><b>if</b> <math>(list \subseteq \text{leafs}(H(BC_{bc}, S_t), t_t))</math> //contrôle qu'au moins un descendant de la nouvelle mesure soit présent dans le Cube de Base <math>BC_{bc}</math></p> <p style="padding-left: 2em;"><b>then</b> <math>\delta'(t_1, \dots, t_{m+1}) := \mathbf{true}</math></p> <p style="padding-left: 2em;"><b>else</b> <math>\delta'(t_1, \dots, t_{m+1}) := \mathbf{false}</math></p> <p style="padding-left: 2em;">}</p> <p><b>Return</b> <math>(\delta')</math></p>

### Exemple 22.

Supposons que l'on applique l'opérateur Specialize à un Vue qui a comme mesures les unités pour répondre à la requête « Pour chaque polluant et pour chaque jour, quelles sont les zones de la lagune telles que au moins une de leurs unités est polluée ? ». Formellement : Soit  $\Theta_{\text{fusion-zone}} = \langle S'_{\text{zone}}, S'_{\text{zone\_fusion}}, \Phi \rangle$ , alors  $\Lambda(V_{\text{corila-unit-1000}})[S_{\text{zone}}, \Theta_{\text{fusion-zone}}] = V_{\text{corila-zone-1000s}} = \langle BC_{\text{corila-zone-1000s}}, \langle S_{\text{pollutant}}, S_{\text{days}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-zone}}, \gamma' \rangle$  où:

- $BC_{\text{corila-zone-1000s}} = \langle H_{\text{pollutants}}, H_{\text{time}}, H'_{\text{lagoon\_spatial}}, H_{\text{rate}} \rangle, \delta' \rangle$  où :
- $\delta' : I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S'_{\text{zone}}) \times I(S_{\text{rate5}})$
- $H'_{\text{lagoon\_spatial}}$  est la Hiérarchie Réduite de  $H_{\text{lagoon\_spatial}}$  sur  $S_{\text{zone}}$
- $\gamma' : I(S_{\text{pollutant}}) \times I(S_{\text{day}}) \times I(S_{\text{rate5}}) \times I(S'_{\text{zone\_fusion}})$

Les tuples instances de  $V_{\text{corila-zone-1000s}}$  sont représentés dans le Tableau 5.10.

$S_{\text{pollutant.name\_pollutant}}$	$S_{\text{day.day}}$	$S'_{\text{zone.name\_zone}}$	$S_{\text{rate5.value5}}$
Zinc	19-9-05	North Swam	45-49
Sulfur Trioxide	19-9-05	North Swam	30-34
Zinc	19-9-05	Bocca Lido	30-34
Benzo(b)fluoronthene	19-9-05	North Swam	45-49
Methane	19-9-05	North Swam	30-34
Zinc	19-9-05	South Swam	45-49
Methane	19-9-05	Bocca Lido	30 -34
Methane	26-9-05	South Swam	30-34
Sulfur Trioxide	26-9-05	Bocca Lido	20-24

Tableau 5.10. Instance de  $V_{\text{corila-zone-1000s}}$

Nous notons que les Tableaux 5.9 et 5.10 sont différents car, même si la granularité de la mesure est la même, ils sont les résultats des opérations de Classify et Specialize et donc de deux requêtes différentes :

- Classify : « *Quelles sont les zones, telles que toutes leurs unités sont polluées, pour chaque jour et pour chaque polluant ?* »
- Specialize : « *Pour chaque polluant et pour chaque jour, quelles sont les zones de la lagune telles que au moins une de leurs unités est polluée ?* »

### 5.3.3 Opérateur de permutation mesure et dimension : Permute

L'opérateur *Permute*, noté  $\Pi$ , permet d'invertir une dimension et une mesure. Le niveau bottom de la hiérarchie devient la nouvelle mesure et la hiérarchie associée à la mesure est utilisée comme axe d'analyse.

Soient  $V_v = \langle BC_{bc}, \mathcal{L}, \Theta_k, \gamma \rangle$  un Schéma de Vue et  $\Theta_t = \langle S_{Lt}, S_{at}, \Theta_t \rangle$  un Mode d'Agrégation tel que  $S_{Lt}$  appartient aux Niveaux de la Vue de  $V_v$ , l'opérateur *Permute* est appliqué à  $V_v$  en utilisant  $\Theta_t$ . Le résultat est un autre Schéma de Vue  $V'_v = \langle BC_{bc}, \mathcal{L}', \Theta_t, \gamma' \rangle$ . Le Cube de Base  $BC_{bc}$  ne change pas car, comme nous l'avons montré précédemment, il ne fait pas la distinction entre mesure et dimension. Les Niveaux de la Vue  $\mathcal{L}'$  contiennent la mesure de  $V_v$ , le Mode d'Agrégation de  $V'_v$  définit une nouvelle mesure, et par conséquent la fonction booléenne de  $V'_v$  aussi change.

Formellement :

*Soient*

- $V_v$  une Vue de Base =  $\langle BC_{bc}, L, \Theta_k, \gamma \rangle$
- $\Theta_t = \langle S_{Lb}, S_{ab}, \Theta_t \rangle$  un Mode d'Agrégation

alors  $\Pi(V_v)[\Theta] = V'_v = \langle BC_{bc}, L', \Theta_b, \gamma' \rangle$  où :

- $L' = \langle S'_{L1}, \dots, S'_{Lm} \rangle$  où :
- $\forall i \in [1 \dots t[ \cup ]t \dots m]$   $S'_{Li} = S_{Li}$
- $S'_{Li} = S_f$  où  $S_f$  est la mesure détaillée du Mode d'Agrégation  $\Theta_k$  de  $V_v$
- $\gamma' : I(S'_{L1}) \times \dots \times I(S'_{Lm}) \times I(S_{at})$ .

### Exemple 23.

Supposons que l'utilisateur veuille introduire les polluants comme mesures à la place des unités, alors il applique l'opérateur Permute de la manière suivante :

$$\Pi(V_{\text{corila-unit}})[\Theta_{\text{list-pollutant}}] = V_{\text{corila-pollutant}}$$

L'instance de  $V_{\text{corila-pollutant}}$  est la même que celle de  $V_{\text{corila-unit}}$  (Tableau 5.1). La représentation graphique de  $V_{\text{corila-pollutant}}$  est montrée en Figure 5.10.

### 5.3.4 Opérateur de modification dynamique de l'hypercube : OLAP-Buffer

Au chapitre 4 (cf chap 4 § 4.1) nous avons montré l'intérêt d'introduire de nouveaux opérateurs multidimensionnels qui modifient la structure de l'hypercube en créant de nouveaux membres à la volée pour adapter les opérateurs d'analyse spatiale, comme le buffer, le clipping, l'overlay, etc. au paradigme multidimensionnel. Parmi ces opérateurs nous nous sommes intéressés dans ce travail à l'opérateur de buffer, n'excluant pas la possibilité d'étendre cette approche aux autres opérateurs.

Pour déterminer par exemple le nombre de personnes qui habitent à moins de 100 mètres de la frontière d'un pays, on détermine une zone dite zone-tampon ou « buffer zone » (Figure 5.16), celle-ci pouvant être à l'intérieur ou à l'extérieur. Pour déterminer cette zone, on fait appel à la morphologie mathématique (Serra, 1969) grâce à un élément structurant et des opérateurs de dilation ou d'érosion. Dans notre cas, l'élément structurant est tout simplement un disque qui « roule » le long de la frontière, et son enveloppe définit ainsi la zone tampon recherchée.



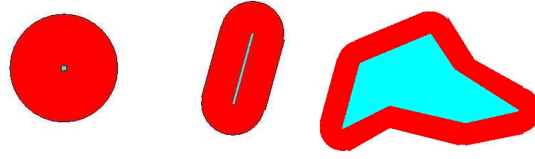


Figure 5.16. Exemples d'application de l'opérateur buffer.

Afin d'intégrer cet opérateur d'analyse spatiale dans GeoCube, nous avons introduit un opérateur, OLAP-Buffer. Il remplace la dimension géographique par une autre dimension géographique qui contient l'objet géographique résultat du buffer. De plus, grâce à des fonctions définies par l'utilisateur qui utilisent les parties des membres recouverts totalement et/ou partialement par la région tampon, les membres originaux et les valeurs de mesures dans la table de faits associés, OLAP-Buffer recalcule les valeurs des mesures pour le nouveau membre géographique.

Formellement :

Soient :

- $V_v$  une Vue de Base =  $\langle BC_{bc}, L, \Theta_k, \gamma \rangle$
- Un Schéma de Hiérarchie  $H_{buffer}$  qui est le même que  $H_{bt}$  sauf pour le niveau bottom qui contient les membres de  $L_{bt}$  et aussi le membre géographique résultat du buffer qui est fils de All.
- $f_{buffer}$  est une fonction qui prend en entrée des membres géographiques et leur mesures associées, et le nouveau membre qui le résultat du buffer et donne comme résultat une mesure

alors,  $\beta(V_v) [H_{buffer}, f_{buffer}] = V'_v = \langle BC'_{bc}, L', \Theta_b, \gamma \rangle$  où :

- $BC'_{bc} = \langle H_{b1'}, \dots, H_{bm+1'}, \delta \rangle$  où :
  - $H_{bt'} = H_{buffer}$  et  $\forall i \in [1 \dots t] \cup [t \dots m+1] H_{bi'} = H_{bi}$
  - $\delta : I(L_{b1'}) \times \dots \times I(L_{bm+1'})$  est calculée en utilisant l'algorithme *BufferBaseCube*
- $L' = \langle S'_{L1}, \dots, S'_{Lm} \rangle$  où :
  - Si  $S_{Lj} = L_{bt}$ , alors  $S'_{Lj} = L_{buffer}$  et  $\forall i \in [1 \dots j] \cup [j \dots m] S'_{Li} = S_{Li}$
- $\gamma : I(S'_{L1}) \times \dots \times I(S'_{Lm}) \times I(S_{af})$  où  $S_{af}$  est la mesure agrégée de  $\Theta_k$ .

**Algorithme BufferBaseCube**

**Entrée**

$V_v$  : Vue de Base à laquelle l'opérateur OLAP-Buffer est appliqué  
 $H_{buffer}$  : Hiérarchie dont le niveau bottom contient le membre qui est le résultat de l'opération de buffer  
 $f_{buffer}$  : fonction utilisé pour le recalcul de la mesure

**Résultat**

$\delta'$  : fonction booléenne du Cube de Base de la nouvelle Vue de Base  $V'_v$

*Soit  $S_f \in L(H_{bf})$  // Le niveau bottom de la hiérarchie  $H_{bf}$  du Cube de Base est utilisée comme mesure*

*Soit  $t_{buffer} \in I(L_{buffer})$  tel que  $t_{buffer}$  est le résultat du buffer*

*Soit  $listD$  un ensemble d'instances d'Entité //  $listD$  va contenir l'ensemble des membres de la dimension géographique*

*Soit  $listM$  un ensemble d'instances d'Entité //  $listM$  va contenir l'ensemble de mesures détaillées associées aux membres géographiques*

**for each**  $\langle t_1, \dots, t_{m+1} \rangle$  **in**  $dom(\delta')$  { // Liste tous les tuples instances possibles de  $V'_v$

**if**  $(\forall i \in [1, \dots, m+1] t_{bi} = t_i)$  **and**  $\delta(t_{b1}, \dots, t_{bm+1}) = \mathbf{true}$  **and**  
 $t_j \neq t_{buffer}$   
**then**  $\delta'(t_1, \dots, t_{m+1}) := \mathbf{true}$  **else**  $\delta'(t_1, \dots, t_{m+1}) := \mathbf{false}$

**if**  $(t_j = t_{buffer})$  {

$listD := \emptyset$

$listM := \emptyset$

**for each**  $\langle t_{b1}, \dots, t_{bm+1} \rangle$  **in**  $dom(\delta)$  { // Liste tous les tuples instances de Cube de Base  $BC_{bc}$

**if**  $(\forall i \in [1, \dots, f-1] \cup [f+1, \dots, t-1] \cup [t+1, \dots, m+1] t_{bi} = t_i)$  **and**  
 $\delta(t_{b1}, \dots, t_{bm+1}) = \mathbf{true}$   
**then** add  $t_{bj}$  to  $listD$  **and** add  $t_{bf}$  to  $listM$

}

**if**  $(t_f = f_{buffer}(t_{buffer}, listD, listM))$  // Contrôle que tous les tuples instances, pour construire la mesure agrégée, sont parmi les instances du Cube de Base

**then**  $\delta'(t_1, \dots, t_{m+1}) := \mathbf{true}$  **else**  $\delta'(t_1, \dots, t_{m+1}) := \mathbf{false}$

}

}

**return**  $(\delta')$

L'idée de l'algorithme BufferBaseCube est de considérer tous les tuples instances  $\langle t_1, \dots, t_{m+1} \rangle$  possibles de  $V_v$  tels que l'objet géographique  $t_i$  n'est pas le résultat de l'opération de buffer ( $t_i$  n'est pas égal à  $t_{buffer}$ ). Alors, s'il existe un tuple  $\langle t_{b1}, \dots, t_{bm+1} \rangle$  de  $V_v$  tel que chaque instance  $t_{bj}$  est égal à  $t_j$ , alors la fonction booléenne du Cube de Base  $BC'_{bc}$  pour  $\langle t_1, \dots, t_{m+1} \rangle$  est vraie.

Ensuite, l'algorithme considère pour tous les autres tuples  $\langle t_1, \dots, t_{m+1} \rangle$ , l'ensemble des tuples  $\{\langle t_{b1}^j, \dots, t_{bm+1}^j \rangle, \text{ où } j=1 \dots n\}$  tels que chaque instance  $t_{bj}^j$ , qui n'est pas une mesure ( $t_{bf}^j$  n'est pas égal à  $t_{bf}$ ) et qui est différente de  $t_{buffer}$ , est égal à  $t_i$ . Si  $t_f$  (la nouvelle mesure) est le résultat de la fonction  $f_{buffer}$  appliquée à  $(t_{bt}^1, t_{bf}^1), \dots, (t_{bt}^n, t_{bf}^n), t_{buffer}$ , alors la fonction booléenne du Cube de Base  $BC'_{bc}$   $\langle t_1, \dots, t_{m+1} \rangle$  est vraie. Nous notons que même si la fonction  $f_{buffer}$  prend en entrée tous les membres géographiques du niveau bottom de la dimension géographique et les mesures associées, elle peut éliminer tous les membres et les mesures dont n'a pas besoin grâce à l'attribut géométrique du membre résultat du buffer.

**Exemple 26.**

Soit  $H_{lagoon\_spatial\_zoneb}$  une hiérarchie égal à  $H_{lagoon\_spatial\_zone}$  sauf pour son niveau bottom  $S_{zoneb}$  qui contient le résultat d'une opération de buffer sur la zone Bocca Malmocco (Figure 5.17). Les représentations graphiques de  $H_{lagoon\_spatial\_zone}$  (Schéma et Instance) et  $H_{lagoon\_spatial\_zoneb}$  (Schéma et Instance) sont montrées en Figure 5.18.

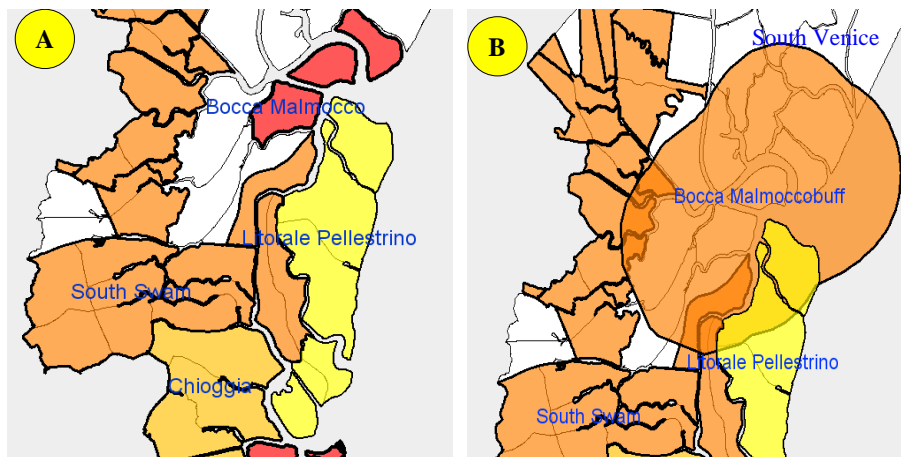


Figure 5.17. Représentation cartographique a)  $S_{zone}$  b)  $S_{zoneb}$

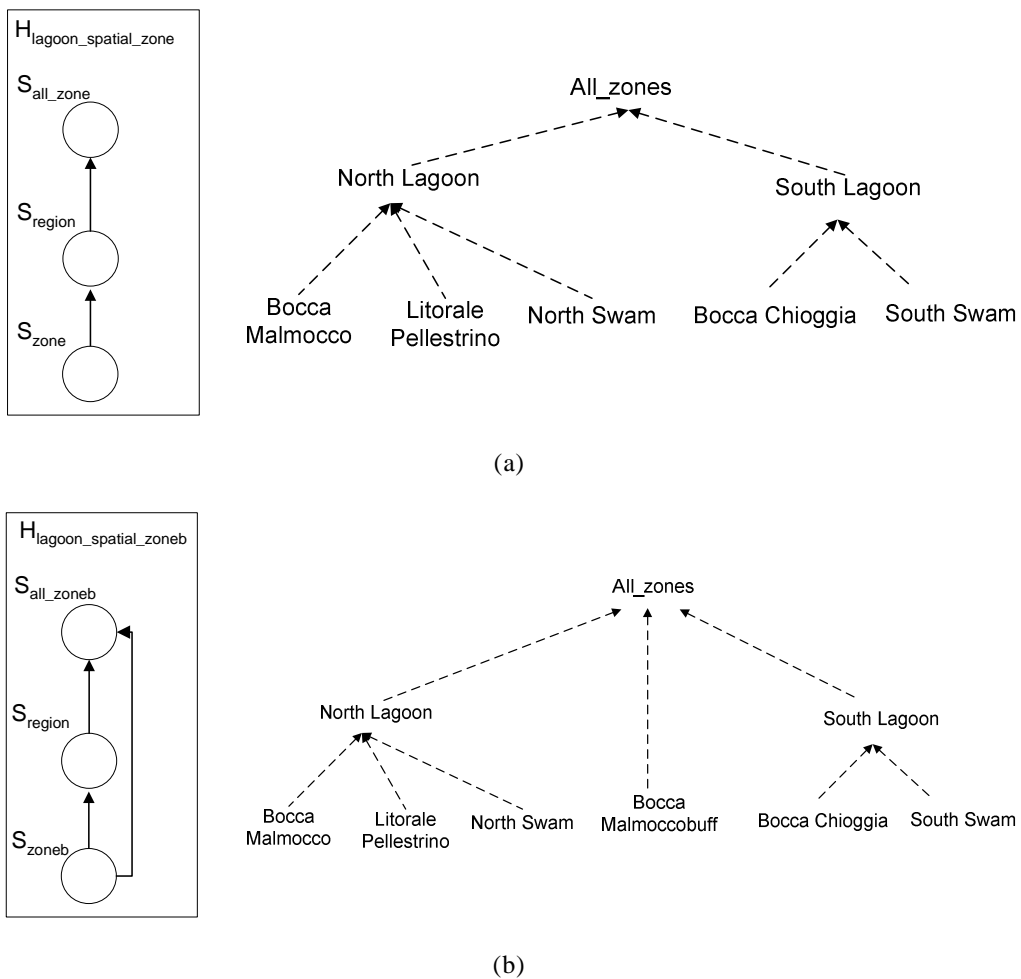


Figure 5.18. Représentation graphique a)  $H_{lagoon\_spatial\_zone}$  b)  $H_{lagoon\_spatial\_zoneb}$

Soit  $V_{rate}$  le Schéma de Vue qui analyse la pollution de la lagune par rapport aux zones, au temps, et aux polluants.

$$V_{rate} = \langle BC_{corila-rate}, \langle S_{pollutant}, S_{day}, S_{zone} \rangle, \Theta_{avg-rate}, \gamma \rangle \text{ où } BC_{corila-rate} = \langle H_{pollutant}, H_{time}, H_{lagoon\_spatial\_zone}, H_{value}, \delta' \rangle \text{ et } \delta' \text{ est définie sur } I(S_{pollutant}) \times I(S_{day}) \times I(S_{zone}) \times I(S_{rate})$$

Soient les données de la table de faits représentées dans le Tableau 5.11.

Soit  $f_{avgbuffer}$  la fonction pour recalculer la valeur de l'attribut de la mesure pour Bocca Mamoccobuff.  $f_{avgbuffer}$  est une moyenne pondérée sur la surface qui utilise les géométries des objets géographiques qui recouvrent la géométrie de Bocca Mamoccobuff, et les mesures associées à ces membres.

$S_{pollutant.name\_pollutant}$	$S_{day.day}$	$S_{zone.name\_zone}$	$S_{rate.value}$
Zinc	19-9-05	South Venice	45
Zinc	19-9-05	Bocca Malmocco	45
Zinc	19-9-05	Litorale Pallesstrino	30
Zinc	19-9-05	South Swam	30
Methane	19-9-05	North Swam	30

Tableau 5.11. Instance de  $V_{rate}$ .

Alors  $b(V_{rate}) [H_{lagoon\_spatial\_zoneb}, f_{avgbuffer}] =$

$V_{rateb} = \langle BC_{corila-rateb}, \langle S_{pollutant}, S_{day}, S_{zoneb} \rangle, \Theta_{avg-rate}, \gamma \rangle$  où  $BC_{corila-rateb} = \langle H_{pollutants}, H_{time}, H_{lagoon\_spatial\_zoneb}, H_{rate}, \delta' \rangle$  où  $\delta' : I(S_{pollutant}) \times I(S_{day}) \times I(S_{zoneb}) \times I(S_{rate5})$

Les tuples instances de  $V_{rateb}$  sont représentés dans le Tableau 5.12.

Pour calculer  $I(V_{rateb})$  l'algorithme BufferBaseCube agit de la façon suivante : pour le tuple  $\langle zinc, 19-9-05, Bocca Malmoccobuff, 42 \rangle$ , l'algorithme cherche dans la table de faits les tuples  $\langle zinc, 19-9-05, South Venice, 45 \rangle$ ,  $\langle zinc, 19-9-05, Bocca Malmocco, 45 \rangle$ ,  $\langle zinc, 19-9-05, Litorale Pallesstrino, 45 \rangle$  et  $\langle zinc, 19-9-05, South Swam, 45 \rangle$ . Puisque 42 est le résultat de l'application de  $f_{avgbuffer}$  sur  $(Bocca Malmocco, 45)$ ,  $(South Venice, 45)$ ,  $(Litorale Pallesstrino, 30)$ ,  $(South Swam, 30)$  et  $Bocca Malmoccobuff$ , alors pour le tuple  $\langle zinc, 19-9-05, Bocca Malmoccobuff, 42 \rangle$  la fonction booléenne est vraie. Ensuite l'algorithme assigne vrai aux tuples  $\langle zinc, 19-9-05, South Venice, 45 \rangle$ ,  $\langle zinc, 19-9-05, Litorale Pallesstrino, 30 \rangle$ ,  $\langle zinc, 19-9-05, South Swam, 30 \rangle$ ,  $\langle zinc, 19-9-05, North Swam, 30 \rangle$  et  $\langle zinc, 19-9-05, Bocca Malmocco, 45 \rangle$ .

$S_{pollutant.name\_pollutant}$	$S_{day.day}$	$S_{zoneb.name\_zoneb}$	$S_{rate.value}$
Zinc	19-9-05	South Venice	45
Zinc	19-9-05	Bocca Malmoccobuff	42
Zinc	19-9-05	Litorale Pallesstrino	30
Zinc	19-9-05	South Swam	30
Methane	19-9-05	North Swam	30
Zinc	19-9-05	Bocca Malmocco	45

Tableau 5.12. Instances de  $V_{rateb}$

L'opérateur OLAP-Buffer restructure l'application multidimensionnelle en ajoutant un membre géographique qui est le résultat d'une opération de buffer. Pour ce nouveau membre la mesure (les attributs) est calculée grâce à une fonction définie par l'utilisateur.

## 5.4 Requêtes multidimensionnelles complexes

Dans la section 4 nous avons présenté les opérateurs de GeoCube. En particulier nous avons introduit les opérateurs Roll-Up, Classify et Specialize. Le Roll-Up permet d'agréger les mesures avec un Mode d'Agrégation, en montant dans une dimension. Les opérateurs Classify et Specialize changent la granularité des mesures en montant dans leur hiérarchie. Dans cette section nous décrivons comment, grâce à la combinaison des opérateurs de l'algèbre, il est possible de changer la granularité des mesure en montant dans une dimension, ou en d'autres termes comment remplacer les processus d'agrégation défini par le Mode d'Agrégation par le changement de la granularité de la mesure.

De la même façon, nous montrons comment il est possible de répondre à des requêtes portant sur des membres cousins dans la hiérarchie de dimension, en couplant l'utilisation des opérateurs Slice et Roll-Up.

### 5.4.1 Reformulation du processus d'agrégation par le changement de la granularité de la mesure

Les opérateurs qui changent la granularité de la mesure ne changent pas les niveaux de dimensions du cuboïde représenté par la Vue, mais ils changent les valeurs des cellules du cuboïde. Grâce aux opérateurs Permute et Specialize, il est possible de redéfinir le processus d'agrégation classique, associé au changement d'un niveau de dimension, avec un changement de granularité. Supposons par exemple que l'utilisateur soit intéressé par les zones de la lagune polluées, pour chaque polluant, pour chaque valeur de pollution et pour chaque mois et que dans la table de faits l'information géographique soit stockée au niveau « Unit ». Pour répondre à cette question il ne peut pas utiliser seulement l'opérateur Roll-Up pour monter au niveau mois, car le Roll-Up agrégera les unités en utilisant le Mode d'Agrégation et donc les mesures ne seront pas des zones de la lagune mais des unités fusionnées. L'idée est alors d'utiliser une séquence d'opérateurs de l'algèbre.

Tout d'abord grâce à l'opérateur Permute, la dimension sur laquelle on veut monter est introduite comme mesure. Ensuite l'opérateur de Specialize est appliqué à la nouvelle mesure. Enfin, une opération de Permute rechange mesure et dimension. Cette séquence d'opérations nous permet de changer la granularité de la dimension, sans agréger les mesures. La dernière opération est alors une opération de Classify ou de Specialize pour changer la granularité de la mesure. Cette séquence d'opérateur peut être vue comme un opérateur composé.

Si l'on considère notre exemple, la séquence des opérateurs que l'on applique sera la suivante :

Le Permute introduit les jours comme mesures. Ensuite une opération de Specialize sur les jours remplacera chaque jour par son mois. Grâce à l'opérateur Permute les mois seront ensuite utilisés en dimension et les unités comme mesures. L'hypercube résultant de ces opérations permettra d'analyser les unités de la lagune par polluant, valeur de pollution et par mois. Nous avons changé la granularité de la dimension temporelle, en passant des jours au mois sans appliquer l'agrégation aux unités. Maintenant à travers l'opérateur de Classify nous changeons la granularité de la mesure en remplaçant les unités par les zones.

L'hypercube obtenu permettra d'analyser alors les zones de la lagune par polluant, par valeur de pollution et par mois.

**Exemple 27.**

Présentons formellement la séquence de ces opérations. Soit  $V_{\text{corila-day}} = \langle BC_{\text{corila}}, \langle S_{\text{pollutant}}, S_{\text{unit}}, S_{\text{rate5}} \rangle, \Theta_{\text{first-day}}, \gamma \rangle$  une Vue de Base qui a comme mesures les jours. Son instance est montrée dans le Tableau 5.1. Appliquons l'opérateur Specialize de la façon suivante :  $\omega(V_{\text{corila-day}})[\Theta_{\text{first-month}}] = V_{\text{corila-month}} = \langle BC_{\text{corila-month}}, \langle S_{\text{pollutant}}, S_{\text{unit}}, S_{\text{rate5}} \rangle, \Theta_{\text{first-month}}, \gamma' \rangle$  où :

- $BC_{\text{corila-month}} = \langle H_{\text{pollutants}}, H'_{\text{time}}, H_{\text{lagoon\_spatial}}, H_{\text{rate}} \rangle, \delta' \rangle$  où :
- $\delta' : I(S_{\text{pollutant}}) \times I(S'_{\text{month}}) \times I(S_{\text{unit}}) \times I(S_{\text{rate5}})$
- $H'_{\text{time}}$  est la Hiérarchie Réduite de  $H_{\text{time}}$  sur  $S_{\text{month}}$
- $\gamma' : I(S_{\text{pollutant}}) \times I(S_{\text{unit}}) \times I(S_{\text{rate5}}) \times I(S_{\text{month\_first}})$

L'instance de  $V_{\text{corila-month}}$  est montrée dans le Tableau 5.13.

Ensuite, nous appliquons ainsi le Permute à  $V_{\text{corila-month}}$  :

$\Pi(V_{\text{corila-month}})[\Theta_{\text{fusion-unit}}] = V_{\text{corila-month-unit}} = \langle BC_{\text{corila-month}}, \langle S_{\text{pollutant}}, S'_{\text{month}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-unit}}, \gamma' \rangle$  où  $\gamma'$  est définie sur  $I(S_{\text{pollutant}}) \times I(S'_{\text{month}}) \times I(S_{\text{rate5}}) \times I(S_{\text{unit\_fusion}})$

Enfin, le Classify est appliqué à  $V_{\text{corila-month-unit}}$  de cette manière :

$\omega(V_{\text{corila-month-unit}})[\Theta_{\text{fusion-zone}}] = V_{\text{corila-month-zone}} = \langle BC_{\text{corila-zone}}, \langle S_{\text{pollutant}}, S'_{\text{month}}, S_{\text{rate5}} \rangle, \Theta_{\text{fusion-zone}}, \gamma' \rangle$  où :

- $BC_{\text{corila-zone}} = \langle H_{\text{pollutant}}, H'_{\text{time}}, H'_{\text{lagoon\_spatial}}, H_{\text{rate}} \rangle, \delta' \rangle$  où :
- $\delta' : I(S_{\text{pollutant}}) \times I(S'_{\text{month}}) \times I(S'_{\text{zone}}) \times I(S_{\text{rate5}})$
- $H'_{\text{lagoon\_spatial}}$  est la Hiérarchie Réduite de  $H_{\text{lagoon\_spatial}}$  sur  $S_{\text{zone}}$
- $\gamma' : I(S_{\text{pollutant}}) \times I(S'_{\text{month}}) \times I(S_{\text{rate5}}) \times I(S_{\text{zone\_fusion}})$

L'instance de  $V_{\text{corila-month-zone}}$  est montrée dans le Tableau 5.14.

$S_{pollutant.name\_pollutant}$	$S'_{month.month}$	$S_{unit.name\_unit}$	$S_{rate5.value5}$
Zinc	9-05	Mazzorbo	45-49
Sulfure Trioxide	9-05	Mazzorbo	30-34
Methane	9-05	Mazzorbo	30-34
Zinc	9-05	Canal Fondello	30-34
Benzo(b)fluoronthé	9-05	Murano	45-49
Methane	9-05	Murano	30-34
Zinc	9-05	Tessera	45-49
Methane	9-05	Coa di Latte	30 -34
Zinc	9-05	Murano	45-49
Sulfure Trioxide	9-05	Coa di Latte	20-24
Methane	9-05	Tessera	30-34
Sulfure Trioxide	9-05	Canal Fondello	20-24

Tableau 5.13. Instance de  $V_{corila-month}$ .

$S_{pollutant.name\_pollutant}$	$S'_{month.month}$	$S'_{zone.name\_zone}$	$S_{rate5.value5}$
Zinc	9-05	North Swam	45-49
Methane	9-05	North Swam	30-34
Sulfure Trioxide	9-05	Bocca Lido	20-24

Tableau 5.14. Instance de  $V_{corila-month-zone}$ .

Nous notons qu'avec cette séquence d'opérations, nous sommes passé du niveau jour au niveau mois sur la dimension temporelle et nous avons changé la granularité de la mesure en remplaçant les unités par les zones.

#### 5.4.2 Roll-up sur des cousins

En OLAP classique, un roll-up permet d'agréger les mesures en remplaçant un ensemble de membres fils par leur père dans la hiérarchie de dimension. Cela signifie que les hiérarchies sont figées. Imaginons une application de géomarketing. Le décideur analyse son chiffre d'affaire en fonction d'un découpage géographique de ses magasins en zones commerciales. Pouvoir modifier à la volée les zones commerciales, autrement dit, regrouper différemment ses magasins, va lui permettre d'optimiser son zonage. Pour cela, il doit pouvoir agréger ses magasins entre cousins. Une variante serait d'agréger seulement une partie de la fratrie.

Dans notre cas d'étude, par exemple, une requête possible peut être « *Quelle est la valeur maximale de pollution pour chaque jour et pour chaque polluant dans la région géographique formée par les unités Mazzorbo et Canal Fondello?* ». Cette requête implique l'application de la fonction d'agrégation (max) sur une région, composée par deux unités (Mazzorbo et Canal Fondello), qui est un membre géographique qui n'appartient pas



à la hiérarchie spatiale (unités, zones) décrivant la lagune de Venise, autrement dit, il s'agit d'une agrégation sur des cousins (Figure 5.19).

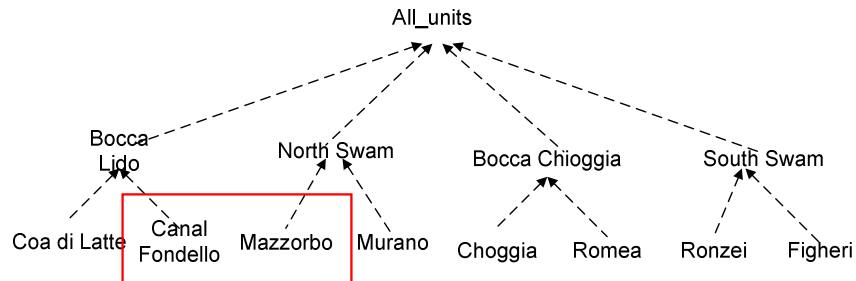


Figure 5.19. Agrégation sur les cousins : Mazzorbo et Canal Fondello.

Notre modèle, grâce aux opérateurs Slice et Roll-Up, permet de répondre à ce type de requête. Nous appliquons le Slice à la dimension géographique en coupant la table de faits sur les membres détaillés sur lesquels nous voulons appliquer la fonction d'agrégation. Ensuite, nous appliquons le Roll-Up en utilisant le niveau top de la même dimension. Cette opération déclenche une opération d'agrégation qui utilise la table de faits coupée par le Slice, et qui donc agrège seulement les mesures associées aux membres qui satisfont le prédicat du Slice.

Dans notre exemple, nous appliquons le Slice sur les membres Mazzorbo et Canal Fondello (Tableau 5.15). Ensuite, nous appliquons le Roll-Up pour monter au niveau top de la dimension géographique (Tableau 5.16). Cette opération associe dans la table de faits le membre « All\_units » à la région géographique composée par les unités Mazzorbo et Canal Fondello. Le Tableau 5.16 nous montre que, le 19/9/05, la valeur maximale de pollution pour le zinc, pour la région géographique composée par les unités Mazzorbo et Canal Fondello est 45-49 mg/l car 45-49 est la valeur maximale entre 45-49 et 30-34.

S <sub>pollutant.name_pollutant</sub>	S <sub>day.day</sub>	S <sub>unit.name_unit</sub>	S <sub>rates.value5</sub>
Zinc	19-9-05	Mazzorbo	45-49
Sulfure Trioxide	19-9-05	Mazzorbo	30-34
Methane	19-9-05	Mazzorbo	30-34
Zinc	19-9-05	Canal Fondello	30-34
Sulfure Trioxide	26-9-05	Canal Fondello	20-24

Tableau 5.15. Slice sur Mazzorbo et Canal Fondello.

$S_{pollutant.name\_polluant}$	$S_{day.day}$	$S_{All\_unit.name}$	$S_{rate5.value5}$
Zinc	19-9-05	All_units	45-49
Sulfure Trioxide	19-9-05	All_units	30-34
Methane	19-9-05	All_units	30-34
Sulfure Trioxide	26-9-05	All_units	20-24

Tableau 5.16. Agrégation entre cousins.

Finalement, GeoCube permet, à travers ses opérateurs, d'agréger sur des ensembles de membres qui ne forment pas de membres de la dimension géographique.

## 5.5 Bilan et comparaison

Dans cette section, nous résumons tout d'abord les caractéristiques principales de notre solution et ensuite nous détaillons les différences principales entre GeoCube et les modèles formels SOLAP introduits au chapitre 4.

### 5.5.1 Caractéristiques principales de GeoCube

GeoCube étend les caractéristiques de modélisation multidimensionnelle que nous avons présentées au chapitre 1, pour prendre en compte des aspects de l'OLAP Géographique. Par rapport aux dimensions, GeoCube gère :

- *Les attributs des dimensions* : Un polluant est décrit par plusieurs attributs par exemple la densité, son code, etc.
- *Les hiérarchies non-strictes* : L'opération de généralisation typification par exemple peut générer une relation plusieurs à plusieurs entre les membres de deux niveaux d'une hiérarchie de généralisation.
- *Les hiérarchies non-couvrantes* : Le zinc n'est pas classé comme aliphatique ou aromatique et par conséquent il n'a pas de correspondant au niveau représentant le type de lien atomique, contrairement, par exemple, au méthane.

GeoCube, pour autoriser la modélisation des mesures de type complexe et géographique gère :

- *Ensemble de mesures* : Une unité de la lagune est décrite par plusieurs attributs comme la géométrie, la salinité, etc.
- *Les mesures multi-valuées* : Plusieurs unités peuvent être polluées par le zinc, le 19-9-05 et avec une valeur de pollution égale à 45-49 mg/l.
- *Les mesures agrégées différentes des mesures détaillées* : L'objet qui représente une mesure agrégée peut présenter plus ou moins d'attributs des mesures détaillées, par exemple l'objet géographique représentant

l'agrégation des unités n'aura pas de nom. De plus, le type de ces attributs peut être différent.

- *Les mesures dérivées* : La surface d'une zone de la lagune est calculée à partir de sa géométrie.
- *Les fonctions d'agrégation définies par l'utilisateur* : Le type d'un ensemble d'unités est calculé comme une moyenne pondérée sur la surface.
- *Mesures au sens classique ayant une signification comme un tout* : La géométrie, le nom, la liste des plantes, la salinité et le type identifient une unité.
- *Les relations plusieurs à plusieurs entre faits et dimensions* : La même unité de lagune peut être polluée le même jour par différents polluants.

Dans notre modèle *une mesure n'est pas prise en compte plusieurs fois dans le processus d'agrégation* : Une unité n'est pas prise en compte plusieurs fois lors d'une opération de Roll-Up.

Une caractéristique fondamentale de GeoCube est la symétrie entre mesures et dimensions. En effet, *les niveaux des dimensions et les mesures sont définis de la même façon* : Une zone de la lagune peut être un niveau de dimension aussi bien qu'une mesure.

Les opérateurs multidimensionnels fournis par GeoCube sont :

- *Opérateurs de navigation dans la hiérarchie de la mesure* : Les unités, même quand elles sont utilisées comme mesures, peuvent appartenir à une hiérarchie (unités, zones). Alors, une requête multidimensionnelle peut porter sur les unités, mais aussi sur les zones de la lagune.
- *Opérateurs de forage et de coupe* : Les zones de la lagune peuvent être analysées selon différentes granularités d'analyse, par exemple jours, mois et années. De plus, l'utilisateur peut réduire l'analyse aux unités de la lagune avec une certaine surface ou analyser la valeur de la pollution seulement en fonction du temps et des polluants.
- *Opérateur qui modifie dynamiquement la structure de l'hypercube* : L'opérateur OLAP-Buffer permet d'analyser la valeur de la pollution par rapport aux polluants, au temps et toutes les unités de la lagune y compris une zone tampon autour d'une unité.
- *Opérateur de permutation mesure et dimension* : L'application multidimensionnelle peut utiliser comme mesures les unités de lagune aussi bien que sur les polluants.

Grâce aux opérateurs multidimensionnels, GeoCube autorise *l'agrégation sur les cousins d'une hiérarchie de dimension* : Une requête multidimensionnelle peut porter sur un ensemble d'unités qui n'est pas incluse dans la dimension géographique.

### 5.5.2 Comparaison

Les modèles formels SOLAP proposés dans la littérature (cf chap 3 § 3), en se focalisant sur des problématiques spatiales particulières, ne prennent pas en compte toutes les caractéristiques de modélisation nécessaires à l'OLAP Géographique. (Ahmed et Miquel, 2005) et (Pourabbas, 2003) s'intéressent respectivement à la modélisation de la dimension spatiale comme un espace continu, et la formalisation de l'intégration d'une base de données multidimensionnelle et d'une base de données géographique. Les deux modèles gèrent seulement les mesures numériques et les hiérarchies non complexes.

Le modèle proposé dans (Jensen, et al., 2004), qui étend celui de Pedersen (Pedersen, 2000), permet de représenter les relations topologiques entre les membres d'une dimension spatiale, et par conséquent d'introduire dans un modèle multidimensionnel les hiérarchies spatiales aussi comme nous l'avons définies au chapitre 4. Le modèle représente une mesure comme une simple valeur, par exemple le nom d'un département. Le modèle est basé sur une symétrie entre dimension et mesure, qui permet d'utiliser tous les niveaux des dimensions comme mesures. En revanche, ce modèle n'autorise pas la définition des mesures dérivées au sens classique et des attributs de dimension, comme souligné en (Abello et al., 2006), car toutes les données de l'application doivent être représentées en tant que niveaux des dimensions. Par conséquent, le modèle ne permet pas de modéliser les mesures de type complexe ou géographique, car un fait est décrit par un ensemble de mesures, mais aussi par tous les niveaux des dimensions.

(Damiani et Spaccapietra, 2006) exploitent la multigranularité de l'information géographique pour définir un modèle où tous les niveaux de la dimension spatiale peuvent être utilisés comme mesures. Le modèle considère une mesure comme un objet géographique, même si les attributs alphanumériques et géométriques ne sont pas explicitement représentés. Le modèle fournit une vision symétrique des mesures et des dimensions car il utilise la même modélisation pour les mesures et pour les dimensions (une mesure spatiale appartient à une hiérarchie comme un niveau de dimension). En revanche, les mesures restent toujours figées dans le processus de navigation comme sujets de l'application multidimensionnelle et il n'y a pas d'opérateur pour intervertir dimension et mesure. Le modèle définit un hypercube pour chaque niveau de la dimension spatiale. L'appartenance de la mesure à une hiérarchie est exploitée grâce à un opérateur qui permet de naviguer parmi ces hypercubes. Ceux-ci doivent être définis a priori car pour sélectionner les nouvelles mesures, il n'est pas possible d'utiliser un prédicat qui utilise les liens hiérarchiques entre les différentes mesures.

Enfin, la modélisation des mesures de type objets complexes est une caractéristique de plusieurs modèles classiques, comme les modèles basés sur le paradigme à objets ((Truijillo et al., 2000), (Nguyen et al., 2000), (Abello, 2006), etc.). En revanche, aucun de ces modèles (Blaschka et al., 1998), (Pedersen, 2000), (Abello et al., 2006) ne présente les mêmes caractéristiques que celles de notre modèle GeoCube. En effet, GeoCube, étant conçu spécialement pour introduire l'information géographique dans un contexte

multidimensionnel, présent des caractéristiques particulières, comme les opérateurs de modification dynamique de l'hypercube, qui sont spécifiques à l'OLAP Géographique.

### 5.5.3 Limites

Plusieurs limites doivent être dépassées pour la prise en compte de l'information géographique dans GeoCube.

Tout d'abord, GeoCube gère seulement « partiellement » les hiérarchies spatiales et de généralisation. La modélisation des relations topologiques d'intersection entre différents membres géographiques pour les hiérarchies spatiales reste un défi majeur. En fait, dans notre solution les fonctions d'agrégation du Mode d'Agrégation utilisent seulement les attributs des mesures détaillées pour le calcul des valeurs des attributs des mesures agrégées. Les géométries des membres géographiques, et donc leurs relations topologiques, alors ne peuvent pas être exploitées dans le processus d'agrégation. Les hiérarchies de généralisation aussi ne sont pas gérées totalement par GeoCube. Par exemple, la modélisation des relations de multi-association entre les membres, ou la définition des fonctions d'agrégation, qui tiennent en compte du changement d'échelle, ne sont pas pris en compte.

GeoCube, en évitant de prendre en compte plusieurs fois la même mesure, fournit un support partiel à la gestion de la sémantique de l'agrégation. En fait, dans notre modèle, il n'y a aucun type de contrainte sur le type de fonction d'agrégation applicable aux attributs des objets complexes ou géographiques (cf chap 1 § 2). De plus, la dépendance entre la fonction d'agrégation spatiale et les fonctions d'agrégation appliquées aux attributs alphanumériques, comme nous l'avons présenté au paragraphe 4.3.2.2.1, n'est pas prise en compte.

## 5.6 Conclusions

L'OLAP Géographique redéfinit les concepts du Spatial OLAP, pour intégrer tous les aspects spatiaux et sémantiques de l'information géographique dans un modèle multidimensionnel. Nous définissons un modèle multidimensionnel pour OLAP Géographique, appelé GeoCube, et une algèbre associée. GeoCube modélise toutes les données de l'application multidimensionnelle, membres et mesures comme des objets géographiques ou complexes, en fournissant une vision symétrique des mesures et des dimensions. Une mesure géographique ou complexe, comme un niveau de dimension, peut appartenir à une hiérarchie et tous les niveaux des dimensions peuvent être utilisés comme sujet d'analyse. L'algèbre permet à l'utilisateur d'agréger les mesures géographiques et complexes, de couper l'hypercube, et d'éliminer une dimension d'analyse. GeoCube, en exploitant la symétrie entre dimension et mesure, définit un opérateur pour intervertir dimension et mesure, et deux opérateurs pour changer la granularité de la mesure. De plus, l'algèbre présente un opérateur qui modifie dynamiquement la structure de l'hypercube grâce à un opérateur d'analyse spatiale.



# Chapitre 6 : *GeWolap, une solution web OLAP Géographique*

Les concepts innovateurs de l'OLAP Géographique, comme nous l'avons montré au chapitre 4, soulèvent de nombreuses problématiques, en ce qui concerne la gestion et la visualisation des données multidimensionnelles, et la navigation dans l'hypercube. Au chapitre 5, nous avons introduit GeoCube, un modèle multidimensionnel, qui, se proposant comme une solution formelle pour OLAP Géographique, intègre dans l'analyse multidimensionnelle la composante sémantique et la composante spatiale de l'information géographique et les opérateurs d'analyse spatiale. Dans ce chapitre, nous présentons notre prototype OLAP Géographique *GeWolap* (Bimonte et al., 2006b), (Bimonte et al., 2007) et (Bimonte et al., 2007b) en utilisant les applications multidimensionnelles détaillées aux chapitres 4 et 5. *GeWolap* est une solution web basée sur le modèle GeoCube qui revisite les solutions OLAP-SIG intégrées classiques en ajoutant un support pour les mesures géographiques et complexes, les dimensions géographiques. De plus, GeoCube introduit des opérateurs d'analyse spatiale dans un contexte multidimensionnel. *GeWolap* permet une analyse spatio-multidimensionnelle effective dans un environnement unique et interactif qui synchronise les composantes cartographique et tabulaire, et les affichages graphiques.

Dans la section 1, nous introduisons l'architecture de *GeWolap*. La dimension géographique et les différents types d'opérateurs d'analyse multidimensionnelle associés sont illustrés en section 2. La section 3 montre comment les mesures géographiques et complexes sont prises en compte dans *GeWolap*. Les fonctionnalités SIG fournies par *GeWolap* sont décrites dans la section 4. Nous concluons le chapitre avec une synthèse des caractéristiques principales de *GeWolap* et des limites du prototype par rapport aux différents opérateurs définis par GeoCube, et par rapport aux aspects de visualisation et d'interaction.

## 6.1 Architecture

*GeWolap* est basé sur une architecture à trois niveaux (Figure 6.1) composée par un SGBD spatial dans lequel sont stockées les données spatiales et alphanumériques, un serveur ROLAP pour la gestion des requêtes multidimensionnelles et un client web. Ce dernier est dédié à la formulation de requêtes multidimensionnelles et à la visualisation des résultats. Il intègre un client SIG et un client OLAP. Dans *GeWolap*, les opérateurs multidimensionnels de GeoCube sont accessibles à travers une carte interactive et la table pivot.

Nous distinguons deux différents types d'opérateurs les opérateurs *multidimensionnels non-spatiaux* qui n'utilisent pas la composante spatiale de l'information géographique et les opérateurs *multidimensionnels spatiaux*.

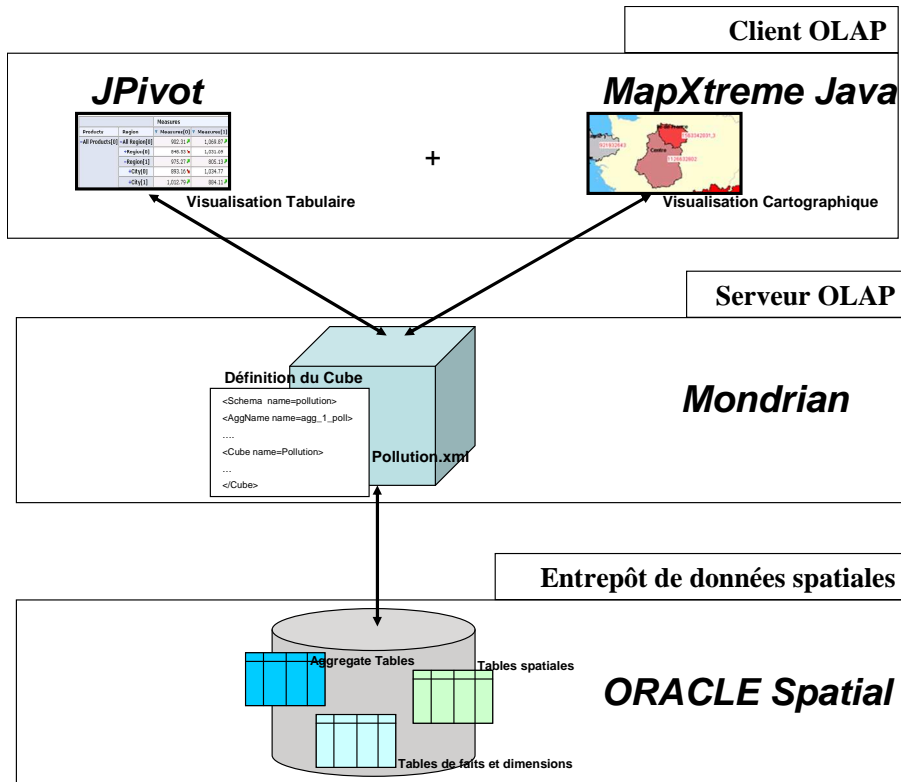


Figure 6.1. Architecture de GeWolap.

Les opérateurs *multidimensionnels spatiaux* utilisent la géométrie des objets géographiques directement ou indirectement. En d'autres termes, les opérateurs *multidimensionnels spatiaux* sont tous les opérateurs multidimensionnels qui utilisent l'interaction avec la carte ou la composante spatiale pour naviguer dans la dimension géographique, pour formuler les prédicats spatiaux ou pour restructurer l'hypercube grâce aux opérateurs d'analyse spatiale. En plus des opérateurs multidimensionnels, GeWolap fournit un ensemble de fonctionnalités SIG. Elles ne déclenchent pas d'opérations de navigation multidimensionnelle, mais elles représentent des fonctionnalités orthogonales aux opérateurs spatio-multidimensionnels.

Puisque Mondrian, comme tous les autres serveurs OLAP existants, commerciaux et non, (Pendse, 2007), (Thomsen et Pedersen, 2005), ne gère pas explicitement la composante spatiale dans les requêtes multidimensionnelles, GeWolap transforme chaque requête multidimensionnelle issue d'un opérateur *multidimensionnel spatial* en une requête classique. Dans un premier temps, GeWolap élabore la composante spatiale, par exemple un prédicat spatial pour la formulation d'une opération de coupe. Ensuite, la requête est reformulée afin d'obtenir une requête multidimensionnelle classique. Ces requêtes sont alors exécutées par Mondrian et Oracle. Les résultats sont visualisés grâce à JPivot. Le



client SIG génère alors, la représentation cartographique du résultat de la requête multidimensionnelle à partir des données spatiales stockées dans Oracle et de la représentation tabulaire de JPivot. Ce processus exige une attention particulière car le pouvoir expressif des cartes et de la table de pivot n'est pas le même, comme nous avons mis en évidence au chapitre 4. Dans le reste de cette section, nous décrivons les fonctionnalités et les caractéristiques principales de chaque niveau.

### 6.1.1 Entrepôt de données spatiales

L'entrepôt de données spatiales est implémenté en utilisant Oracle (Oracle9i, 2007). Les données spatiales et alphanumériques de l'application multidimensionnelle sont stockées dans Oracle en utilisant une modélisation en étoile ou en flocon (cf chap 1 § 3.2.2).

Oracle est un SGBD Objet-relationnel qui fournit un support natif pour les données spatiales (OracleSpatial, 2007). Grâce à Oracle, GeWolap s'enrichit de tous les avantages d'une solution fortement intégrée pour la mémorisation des données spatiales, comme nous l'avons décrit au chapitre 2 : sécurité, accès standard aux données, efficacité et passage à l'échelle. La possibilité de mémoriser une grande quantité de données spatiales et de obtenir de temps de réponse optimisés, représentent des caractéristiques incontournables de l'analyse spatiale en ligne surtout au cas où l'entrepôt de données contient une mesure avec la composante spatiale (Marchand, et al., 2003) et (Newell, 1990). De plus, Oracle offre la possibilité de définir des données complexes, en utilisant des données atomiques, et des fonctions d'agrégation ad hoc. Ces deux fonctionnalités sont nécessaires à la modélisation et à l'agrégation des objets géographiques et complexes. Les attributs des mesures, comme nous l'avons mis en évidence aux chapitres 4 et 5, peuvent être des données complexes et les fonctions d'agrégations utilisées pour les agréger peuvent être différentes des agrégations SQL classiques, par exemple une moyenne pondérée sur la surface.

Enfin, l'entrepôt de données contient aussi un ensemble de métadonnées, comme le nom de l'application et le projet d'appartenance et le type d'application (mesures classique, mesures géographiques, etc.). Il mémorise aussi les informations concernant la dimension qui doit être cartographiée (la dimension géographique), les fonctions d'agrégation pour les attributs des mesures géographiques et complexes, et un ensemble des fonctions pour recalculer les valeurs des attributs des mesures pour les membres calculés à la volée par les opérateurs qui changent dynamiquement l'hypercube grâce aux opérateurs d'analyse spatiale que nous avons introduits aux chapitres 4 et 5.

### 6.1.2 Serveur OLAP

Le serveur OLAP a l'importante tâche de gérer les dimensions et les mesures géographiques et complexes, et d'implémenter les opérateurs de GeoCube. Pour ce niveau, nous avons utilisé Mondrian (Mondrian, 2007). Mondrian est un serveur OLAP « code source libre » développé en Java, qui s'appuie sur une technologie relationnelle. Mondrian traduit toutes les requêtes multidimensionnelles MDX générées par le client OLAP, en requêtes SQL, et il établit un lien entre le client OLAP et le SGBD. Une requête MDX

définit l'ensemble des membres des dimensions et des mesures qui doivent être affichés dans une table de pivot en lignes et en colonnes.

Mondrian fournit toutes les fonctionnalités les plus importantes d'un serveur OLAP, comme les hiérarchies multiples, les propriétés de niveaux des dimensions, les fonctions définies par l'utilisateur, les membres calculés, etc. Le modèle de la base de données multidimensionnelle, appelé schéma, est défini à travers un fichier XML. Il représente aussi le « mapping » du modèle de l'application multidimensionnelle vers les tables stockées dans le SGBD. Un exemple de schéma est montré en Figure 6.2. Il modélise une application multidimensionnelle avec un cube « Sales », représenté par la table « sales\_fact\_1997 » et deux dimensions « Time », avec la hiérarchie jour, mois et années, et « Product » qui est composée par une hiérarchie représentant les produits classés par leur type. Ces deux dimensions sont représentées respectivement par les tables « TimeT » et « ProductT ». Le schéma définit une seule mesure classique « Prod Sales » qui est agrégée en utilisant la somme.

Une requête MDX, portant sur ce hypercube et sa traduction en SQL sont montrées en Figure 6.3. Ces requêtes représentent la somme des produits vendus chaque année.

Mondrian gère seulement les agrégations SQL classiques (SUM, MIN, MAX, COUNT, AVG et DISTINCT COUNT). Nous avons alors étendu ses fonctionnalités pour gérer l'agrégation des objets complexes et géographiques, en ajoutant une nouvelle fonction d'agrégation appelée « UserDefinedAggregation ». Son fonctionnement est décrit en détail dans le paragraphe 6.3.1.

Une fonctionnalité très importante de Mondrian est le support pour les vues matérialisées, qui, comme montré au paragraphe 3.1, est indispensable pour la prise en compte des mesures géographiques et complexes et pour autoriser une vision symétrique des mesures et des dimensions. Une vue matérialisée est implémentée en Mondrian à travers un « Aggregate Table ». Un Aggregate Table est définie dans le schéma de l'application multidimensionnelle de la même façon que tous les autres éléments multidimensionnels. Un exemple d'Aggregate Table est montré en Figure 6.4. Ce code XML représente le cuboïde de l'application de Figure 6.2 avec les niveaux années et produits. Il présente la mesure classique pré-calculée « Prod Sales » et un ensemble de niveaux pour chaque dimension. Les niveaux les plus détaillés « Year » et « Product » identifient les niveaux du cuboïde représenté par l'Aggregate Table.

```

<xml version="1.0" ?>
<Schema>
  <Cube name="Sales">
    <Table name="sales_fact_1997">

      <Dimension name="Time" foreignKey="dayid">
        <Hierarchy hasAll="true" allMemberName="All Time"
          primaryKey="dayid">
          <Table name="TimeT" />
          <Level name="Year" column="yearid"
            uniqueMembers="false" />
          <Level name="Month" column="monthid"
            uniqueMembers="false" />
          <Level name="Day" column="dayid" uniqueMembers="false" />
        </Hierarchy>
      </Dimension>
      <Dimension name="Product" foreignKey="productid">
        <Hierarchy hasAll="true" allMemberName="All Product"
          primaryKey="productid">
          <Table name="ProductT" />
          <Level name="Type" column="typeid" uniqueMembers="false"
            />
          <Level name="Product" column="productid"
            uniqueMembers="false" />
        </Hierarchy>
      </Dimension>
      <Measure name="Prod sales" column="prod_sales"
        aggregator="SUM" />
    </Cube>
  </Schema>

```

Figure 6.2. Fichier XML représentant le schéma Mondrian d'une application multidimensionnelle.

```
SELECT [Measures].[Prod sales] ON COLUMNS,
CROSSJOIN {[Time].[Year].Members}, {[Product].[Food]}
ON ROWS
FROM SALES
```

(a)

```
select
"TimeT"."Year" as "c0",
"ProductT"."Type" as "c1",
sum("sales_fact_1997"."prod_sales") as "m0"
from
"sales_fact_1997" "sales_fact_1997",
"TimeT" " TimeT ",
"ProductT" " ProductT ",
where
"sales_fact_1997"."dayid" = "TimeT"." dayid " and
"sales_fact_1997"."productid" = "ProductT"."productid"
and
"TimeT"."Year" in ('2000', '2001') and
"ProductT"."Type" in ('Food')
group by
"ProductT"."Type",
"TimeT"."Year"
```

(b)

Figure 6.3. Requête multidimensionnelle a) MDX b) Traduction en SQL.

```
AggName name="AGGFACTSSALES_1997">
  <AggFactCount column="FACT_COUNT" />
  <AggMeasure name="[Measures].[Prod Sales]"column="prod_sales"
  />
  <AggLevel name="[Product].[Type]" column="typeid" />
  <AggLevel name="[ Product].[Product]" column="productid" />
  <AggLevel name="[Time].[Year]" column="yearid" />
</AggName>
```

Figure 6.4. Définition d'un Aggregate Table.

### 6.1.3 Client OLAP

Le client est développé en utilisant JPivot et MapXtreme Java. JPivot (JPivot, 2007) est une bibliothèque de balises JSP qui permet d'afficher les résultats des requêtes multidimensionnelles grâce à une table de pivot et à des affichages graphiques. De plus, il permet à l'utilisateur à formuler toutes les opérations de navigation OLAP classiques comme le slice, le dice, le roll-up ou le drill-down à travers la simple interaction avec les composantes visuelles. JPivot peut utiliser Mondrian comme serveur OLAP et des sources de données XMLA. MapXtreme Java (MapInfoMJ, 2007) est une API java pour la réalisation d'applications web cartographiques. MapXtreme Java permet de créer et visualiser des cartes dans tous les navigateurs web standard. Il implémente toutes les fonctionnalités SIG de base, comme l'interaction avec les cartes pour la récupération des

valeurs des attributs alphanumériques des objets géographiques, ou pour déplacer, agrandir et diminuer la taille des cartes. De plus, cette API java permet à l'utilisateur de créer des cartes thématiques, et d'introduire dans l'application web cartographique des simples fonctionnalités d'analyse spatiale.

L'interface de notre prototype représente un environnement interactif qui encapsule la structure de l'application spatio-multidimensionnelle et traduit les interactions avec l'interface en opérateurs SIG, multidimensionnels non-spatiaux et multidimensionnels spatiaux. L'interface combine et synchronise la composante cartographique, la table de pivot et les affichages graphiques. La composante cartographique est une carte interactive qui répond visuellement aux actions de l'utilisateur, en déclenchant des opérateurs multidimensionnels et SIG, et en modifiant les affichages graphiques. Plus précisément, l'utilisateur a accès aux opérateurs de forage et de coupe à travers la simple interaction avec la table de pivot ou la composante cartographique. Les opérateurs qui modifient dynamiquement la structure de l'hypercube et les opérateurs SIG sont accessibles à travers une simple interaction avec la composante cartographique. Les résultats des ces opérateurs sont affichés dans les trois composantes. En particulier, la composante cartographique adapte la visualisation et ses fonctionnalités en fonction du type de mesures : numérique, géographique ou complexe, et du type de dimension : géographique ou classique.

## 6.2 L'interface visuelle de GeWOLap

L'interface est composée de deux panneaux : le *Control Panel* et le *Data Panel* (Figure 6.5). Le *Data Panel* permet de visualiser et de naviguer dans l'hypercube, et le *Control Panel* permet de personnaliser les composantes visuelles du *Data Panel*.

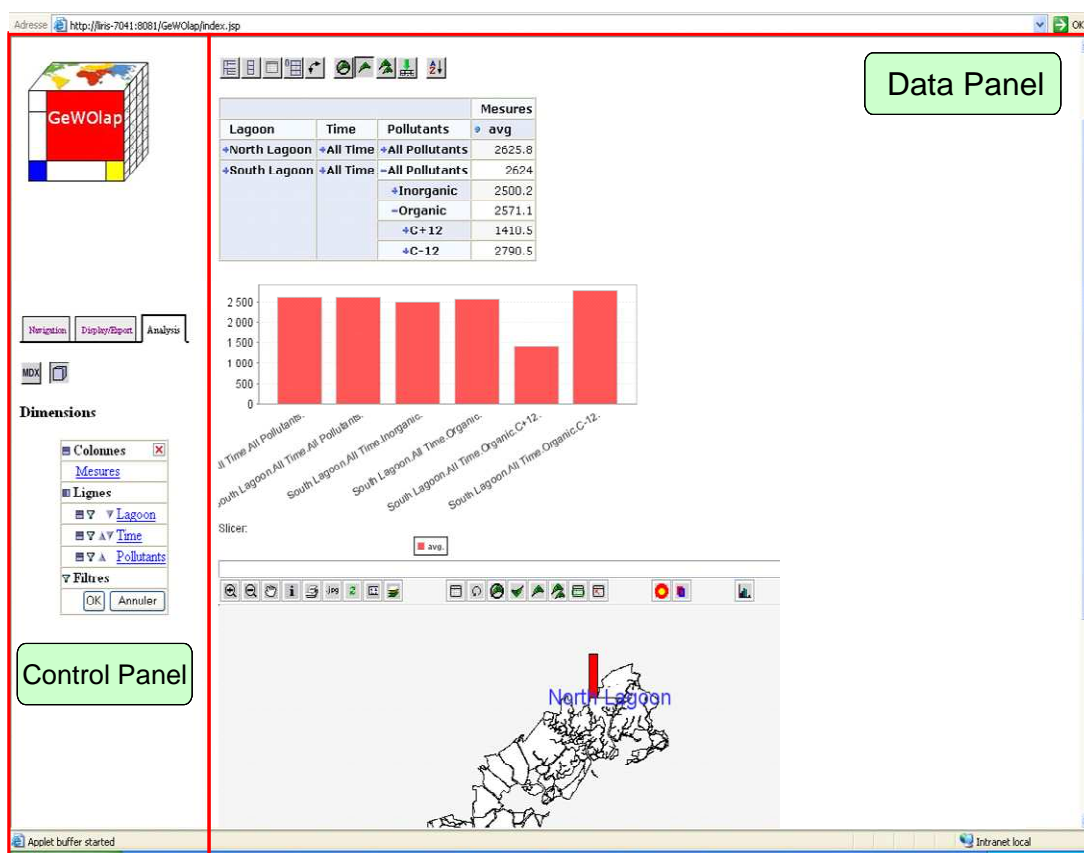


Figure 6.5. Interface visuelle de GeWOLap : Control Panel et Data Panel.

## 6.2.1 Le Data Panel

Le Data Panel combine la table de pivot, les affichages graphiques et une composante cartographique dans un seul environnement. De plus, il présente 3 toolbars qui fournissent l'accès aux différents opérateurs multidimensionnels et SIG.

### 6.2.1.1 Toolbar SIG

La toolbar SIG (Figure 6.6) propose des fonctionnalités purement SIG : *Zoom in* et *Zoom out* qui agrandissent et diminuent respectivement la taille de l'image de la carte, *Pan* qui permet de déplacer la carte, *Info Tool* qui permet d'avoir toutes les informations alphanumériques associées aux objets géographiques, *Map Print* et *Map Export* qui permettent d'imprimer et d'exporter la carte sous forme d'image, *Rule Tool* qui calcule la distance entre deux points et le *Control Layer*. Ceci permet d'ajouter des couches vectorielles et/ou raster, de personnaliser l'affichage de la carte, et de formuler des requêtes sur la base de données spatiales.

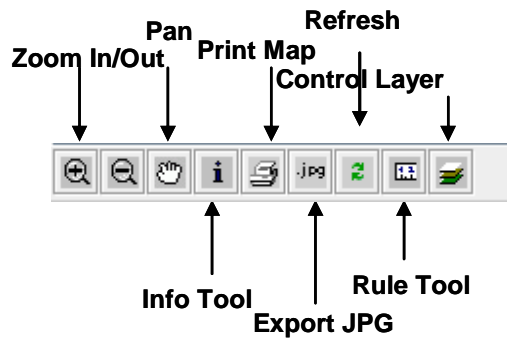


Figure 6.6. Toolbar SIG.

### 6.2.1.2 Toolbar multidimensionnelle non-spatiale

La toolbar multidimensionnelle non-spatiale (Figure 6.7) fournit les fonctionnalités pour modifier le contenu de la table de pivot et choisir parmi les opérateurs multidimensionnels non-spatiaux : *drill-down replace*, *drill-down position*, *expand-all*, et *drill-through*. Le *drill-down replace* autorise une opération de forage d'un membre vers tous ses fils dans une hiérarchie de dimension, en cachant le membre père. Le *drill-down position* garde le membre père, et le *expand-all* permet de forer à partir d'un membre tous les membres fils dans la table.

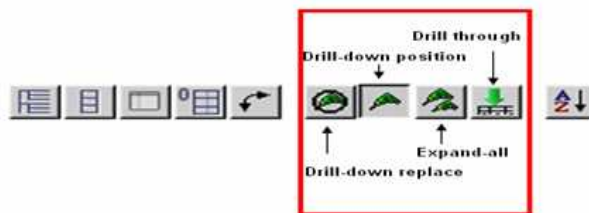


Figure 6.7. Toolbar multidimensionnelle non-spatiale.

### 6.2.1.3 Toolbar multidimensionnelle spatiale

La toolbar multidimensionnelle spatiale (Figure 6.8), qui est disponible seulement en présence d'une application avec une dimension géographique, permet de choisir les opérateurs multidimensionnels spatiaux. Pour chaque opérateur de la toolbar multidimensionnelle non-spatiale, la toolbar multidimensionnelle spatiale met à disposition un opérateur correspondant : *GeWolap drill-down position*, *GeWolap drill-down member* et *GeWolap expand-all*. En supplément, l'opérateur *GeWolap roll-up* permet de monter dans la dimension géographique. En plus du *Cube Browser* du Control Panel, GeWolap présente deux opérateurs pour sélectionner un ensemble de membres de la dimension géographique : le *GeWolap slice position* et *GeWolap slice predicate*. Les opérateurs *GeWolap overlay* et *GeWolap buffer* introduisent dans un contexte multidimensionnel les opérateurs d'analyse spatiale de transformation de buffer et overlay. Enfin, la toolbar multidimensionnelle spatiale fournit une fonctionnalité (*Properties Member*) qui affiche les propriétés d'un membre de la dimension géographique et une autre (*Measure displays*) qui

permet de paramétrer le type de diagramme représentant les mesures classiques, en choisissant entre des barres et des camemberts.

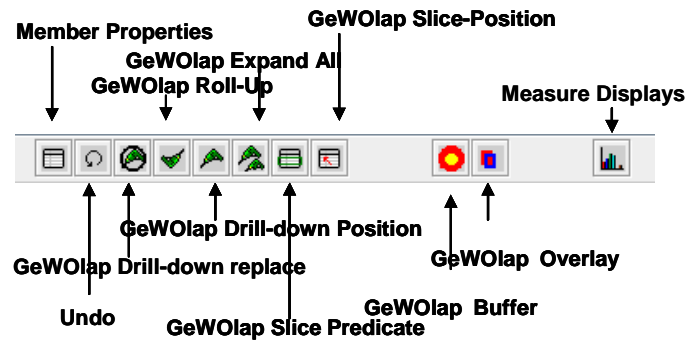


Figure 6.8. Toolbar multidimensionnelle spatiale.

### 6.2.2 Le Control Panel

Le *Control Panel* est divisé en trois onglets :

- L'onglet *Navigation* (Figure 6.9a) liste tous les hypercubes disponibles pour chaque projet. Chaque hypercube présente une icône différente, selon que l'hypercube utilise des mesures numériques, une dimension géographique et des mesures numériques ou une mesure géographique ou une mesure complexe. Par exemple, dans le projet CORILA nous avons défini un hypercube « Corila Pollution » avec une dimension géographique et des mesures numériques, un hypercube avec une mesure géographique « Corila zone », un hypercube sans aucune composante géographique « Quality » et un hypercube avec une dimension géographique et une mesure complexe « Corila polluant ».

- L'onglet *Display/Export* (Figure 6.9b) permet de visualiser/cacher la table de pivot, les affichages graphiques et/ou la carte, et d'exporter la table de pivot vers les formats pdf et excel. Ces fonctionnalités sont très importantes dans un système d'aide à la décision car les données doivent pouvoir être explorées, analysées mais aussi présentées et communiquées.

- L'onglet *Analysis* (Figure 6.9c) contient un navigateur, appelé *Cube Browser*, qui affiche les dimensions, les niveaux et les membres de l'hypercube. Il est utilisé pour sélectionner les membres des dimensions et les mesures qui doivent être visualisées sur les axes de la table de pivot (« Colonnes » et « Lignes »). De plus, le *Cube Browser* permet d'éliminer une dimension d'analyse en sélectionnant un de ses membres à travers le champ « Filtres ».





Figure 6.9. Control Panel a) Navigation b) Display Export c) Analysis.

### 6.3 Dimension Géographique et opérateurs multidimensionnels

Dans cette section, nous détaillons la façon dont GeWolap gère les dimensions géographiques et les opérateurs multidimensionnels (spatiaux et non-spatiaux) associés. Plus précisément, en utilisant l'application multidimensionnelle concernant la pollution de la lagune de Venise montrée en Figure 6.10, nous montrons dans le paragraphe 6.3.1, comment l'interface prend en compte les opérateurs multidimensionnels non-spatiaux de forage et de slice implémentés par JPivot et Mondrian, et dans le paragraphe 6.3.2 nous présentons les opérateurs multidimensionnels spatiaux, en montrant comment les opérateurs de slice et de modification dynamique de l'hypercube, définis par GeoCube, sont implémentés.

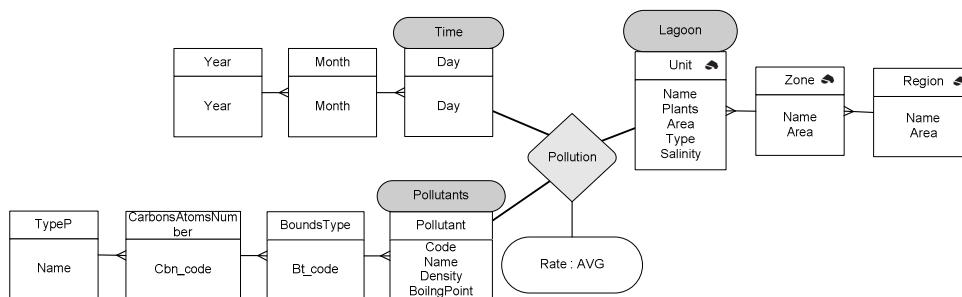


Figure 6.10. Schéma conceptuel d'une application avec une hiérarchie spatiale.

L'application multidimensionnelle de la Figure 6.10 présente une dimension géographique composée par une hiérarchie spatiale avec les niveaux géographiques « Unit », « Zone » et « Region ». La mesure « Rate » est une mesure numérique représentant la valeur de la pollution, qui est agrégée en utilisant la moyenne.

Le fichier XML du schéma Mondrian pour cette application est présenté en Figure 6.11 où la dimension « Lagoon » est définie comme étant la dimension géographique. Un schéma en flocon a été adopté pour la modélisation logique. La dimension géographique est normalisée et les trois niveaux différents correspondent à trois couches ou tables spatiales différentes.

```

<xml version="1.0" ?>
<Schema name="POLLUTION">
  <Cube name="POLLUTION">

    <Table name="POLLUTIONNUMFACTS"/>
    <Dimension name="Lagoon" foreignKey="UNIT_CODE">
      <Hierarchy hasAll="false" primaryKey="UNIT_CODE"
primaryKeyTable="UNIT_CODE">

        <Join leftKey="ZONE_CODE" leftAlias="" rightAlias="ZONE"
rightKey="ZONE_CODE">
          <Table name="UNIT" />
          <Join leftKey="REGION_CODE" leftAlias="" rightAlias=""
rightKey="REGION_CODE">
            <Table name="ZONE" />
            <Table name="REGION" />
            </Join>
          </Join>
          <Level name="Region" table="REGION" column="Region"
uniqueMembers="TRUE" />
          <Level name="Zone" table="ZONE" column="Zone">
          <Level name="Unit" table="UNIT" column="Unit" uniqueMembers="true" />
          </ Hierarchy >
        </Dimension>
    <Dimension name="Time" foreignKey="DAY_CODE">
      <Hierarchy hasAll="true" allMemberName="All Time"
primaryKey="DAY_CODE">
        <Table name="TIMEDIM" />
        <Level name="Year" column="YEAR" uniqueMembers="false"/>
        ...
        <Level name="Day" column="DAY" uniqueMembers="false" />
        </ Hierarchy >
      </Dimension>
    <Dimension name="Pollutant" foreignKey="POLLUTANT_CODE">
      <Hierarchy hasAll="true" allMemberName="All Pollutant"
primaryKey="POLLUTANT_CODE">
        <Table name="POLLUTANTDIM" />
        <Level name="Type" column="TYPE" uniqueMembers="false" />
        ...
        <Level name="Pollutant" column="POLLUTANT"
uniqueMembers="false" />
        </ Hierarchy>
      </Dimension>
    <Measure name="avg" column="VALUE" aggregator="Avg" />
  </Cube>
</Schema>

```

Figure 6.11. Schéma Mondrian de l'application de Figure 6.10.

### 6.3.1 Opérateurs multidimensionnels non-spatiaux

Supposons que l'utilisateur veuille répondre à la question : « *Quelle est la pollution moyenne au zinc, pour chaque région de la lagune et pour tous les jours ?* ».

Pour répondre à cette requête, en utilisant le Cube Navigator, l'utilisateur sélectionne les membres des différentes dimensions qui doivent être affichées : « Zinc » dans la dimension des polluants, « All Time » dans la dimension temporelle et « North Lagoon » et « South Lagoon » dans la dimension géographique, comme l'illustre la Figure 6.12.

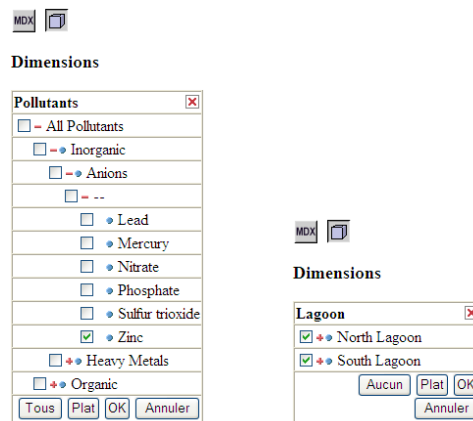


Figure 6.12. Sélection des membres de la dimension des polluants et de la dimension géographique à travers le Cube Navigator.

L'interaction avec le Cube Navigator se traduit par la création d'une requête MDX (Figure 6.13) dont le résultat est visualisé par GeWolap, sous forme tabulaire et cartographique (Figure 6.14).

```
select {[Measures].[avg]} ON COLUMNS,
    Crossjoin({[Lagoon].[North Lagoon],
[Lagoon].[South Lagoon]}, {[Time].[All
Time],
[Pollutants].[AllPollutants].[Inorganic].[
Anions].[Zinc]}) ON ROWS
    from [CORILA_Pollution]
```

Figure 6.13. Requête MDX générée par JPivot à travers l'interaction avec la table de pivot.

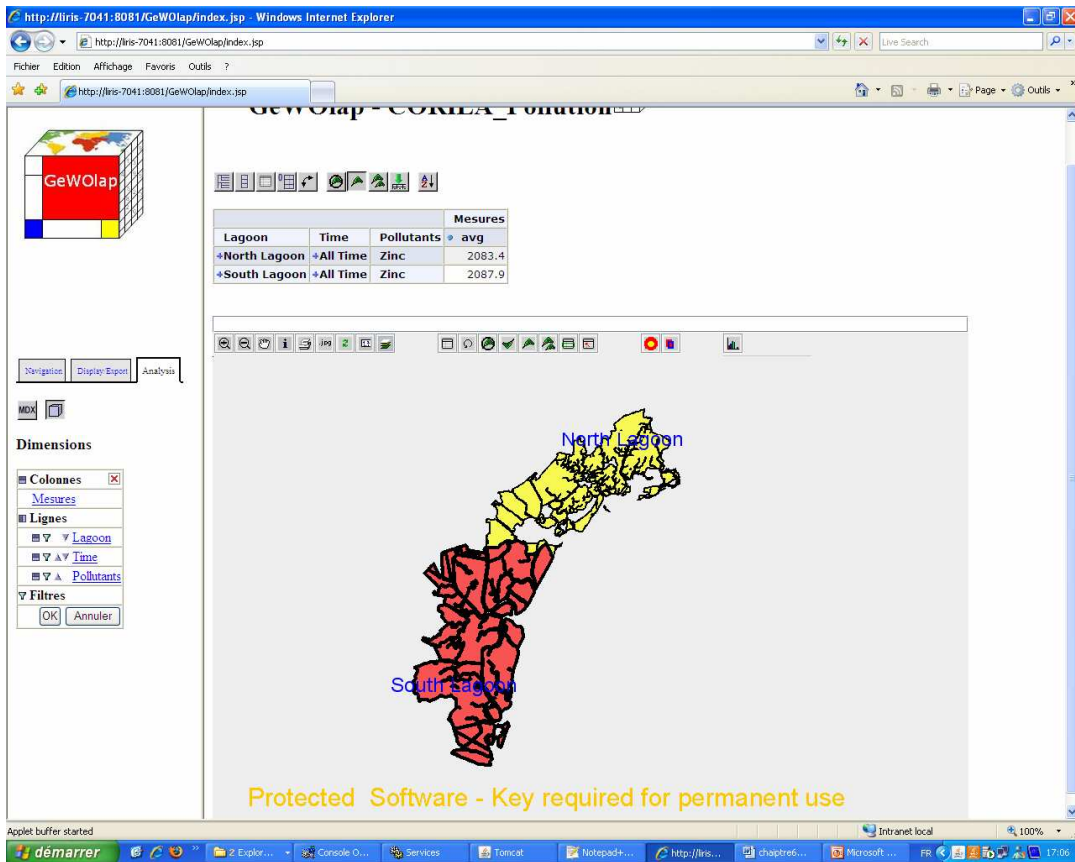


Figure 6.14. Visualisation de la requête multidimensionnelle de Figure 6.13.

La dimension géographique « Lagoon » est visualisée grâce à une carte choroplète interactive où les valeurs de la mesure sont représentées par différentes couleurs.

Supposons que l'utilisateur soit intéressé par les valeurs de la pollution pour toutes les zones de la région « North Lagoon ». Il s'agit d'une opération de forage et plus particulièrement d'une opération de *drill-down position* sur un membre de la dimension géographique. A travers la simple interaction avec la table de pivot, il clique sur la cellule associée au membre « North Lagoon ». Alors, GeWolap exécute cette nouvelle requête multidimensionnelle et reconfigure l'interface pour visualiser le résultat (Figure 6.15). Puisque les différentes composantes sont synchronisées, alors cette opération, qui change le niveau de la dimension géographique, modifie l'affichage tabulaire, et aussi la représentation cartographique. Nous notons que GeWolap, contrairement à l'approche SIG fortement ou faiblement couplée, crée à la volée une carte qui mélange des objets géographiques différents : les zones de « North Lagoon » et la région « South Lagoon ». De plus, même si dans la table de pivot les informations concernant les valeurs de la pollution pour la région « North Lagoon » sont toujours présentes, elles ne peuvent pas être affichées sur la carte car une région recouvre ses zones. Cela nous montre que GeWolap, en prenant en compte les relations topologiques entre les membres de niveaux différents, autorise une complète synchronisation de ses composantes visuelles.

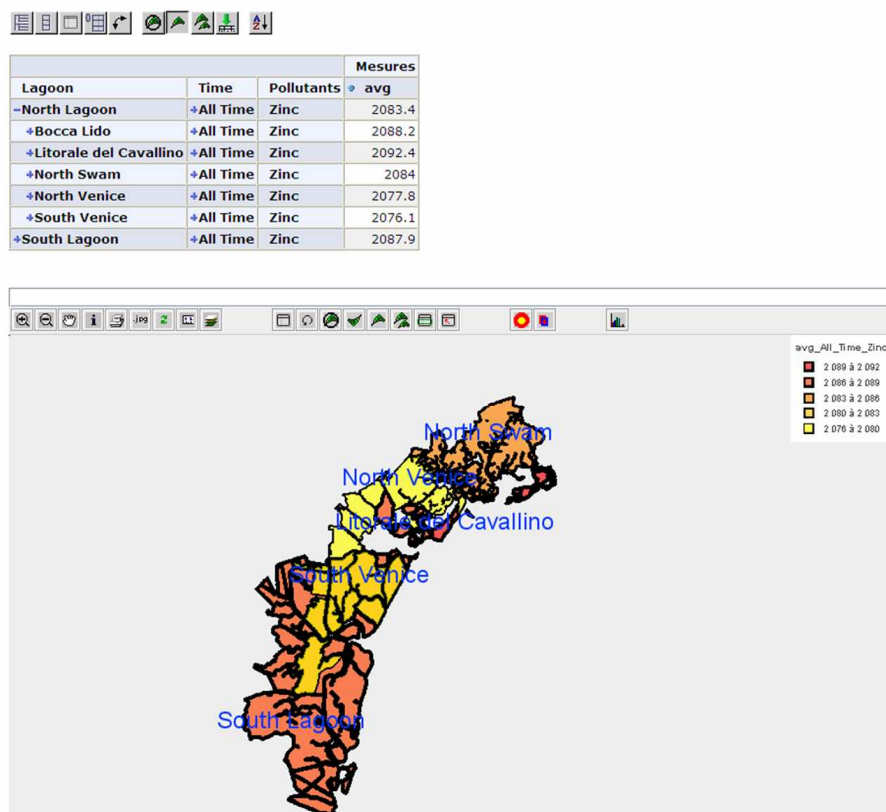


Figure 6.15. Relations topologiques et synchronisation de la composante cartographique et de la table de pivot.

A cette étape du processus de navigation, l'utilisateur décide de découvrir la valeur moyenne de la pollution pour chaque zone et pour les mois de septembre et octobre 2002. Il s'agit d'une opération de forage sur la dimension temporelle et ensuite d'une sélection sur les deux mois. Pour répondre à cette requête, il applique les opérateurs *expand-all* et l'opérateur de coupe sur la dimension « Time ». La Figure 6.16 affiche la visualisation du résultat de cette requête.

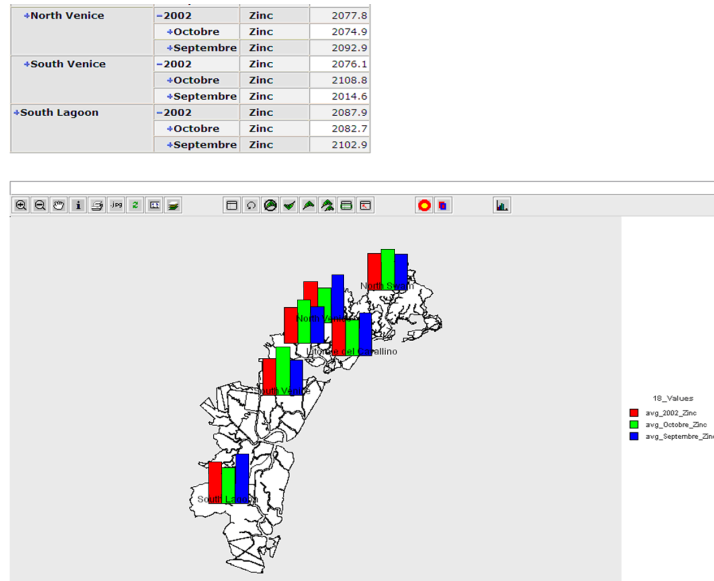


Figure 6.16. Drill-down sur une dimension classique et synchronisation de la composante cartographique et de la table de pivot.

Comme montré en Figure 6.16, GeWolap affiche maintenant une carte avec trois barres. Elles représentent la valeur de la pollution pour le 2002, pour le mois de Septembre et pour le mois d’Octobre. Cette nouvelle configuration visuelle est le résultat d’une opération de forage sur une dimension classique et du mécanisme de synchronisation entre la table de pivot et la carte, qui montrent toujours la même information multidimensionnelle. Nous remarquons comment une opération de forage transforme une carte choroplèthe en une carte thématique avec des barres. Cela signifie que la carte interactive de GeWolap adapte automatiquement la visualisation des mesures classiques selon le nombre de valeurs qui doivent être affichées. Actuellement la règle de visualisation qui utilise une carte choroplèthe pour la visualisation d’une seule mesure classique et des cartes thématiques avec des barres ou des camemberts pour plusieurs mesures classiques est la seule règle de visualisation implémentée dans notre prototype.

Les opérateurs classiques de forage et de coupe de JPivot et Mondrian changent ainsi, outre la configuration visuelle de la table de pivot, l’affichage cartographique de la dimension géographique, afin de synchroniser les trois composantes de l’interface.

### 6.3.2 Opérateurs multidimensionnels spatiaux

Dans cette section, nous décrivons les opérateurs multidimensionnels spatiaux de forage, de coupe et de changement dynamique de l’hypercube.

#### 6.3.2.1 Opérateurs de forage

Les opérateurs de forage *GeWolap roll-up*, *GeWolap drill-down position*, agissent de la même manière que leurs opérateurs multidimensionnels non-spatiaux respectifs. Ils changent le niveau de la dimension géographique et modifient la configuration de la carte,

de la table de pivot et des affichages graphiques, comme nous l'avons décrit précédemment.

Supposons par exemple qu'à partir de la configuration cartographique montrée en Figure 6.17-1, l'utilisateur décide d'appliquer le *GeWolap drill-down position* sur le membre « North Lagoon ». Un simple clic de la souris sur cet objet géographique, déclenche alors l'opération de forage. Le résultat de cette requête est montré en Figure 6.17-2. Nous notons que l'opération de forage modifie la table de pivot et la carte. La carte initiale présente une mesure classique qui représente la pollution moyenne pour tous les polluants pour la région nord de la lagune (Figure 6.17-1). Après le forage, la carte montre la pollution moyenne pour toutes les zones de la région nord (Figure 6.17-2).

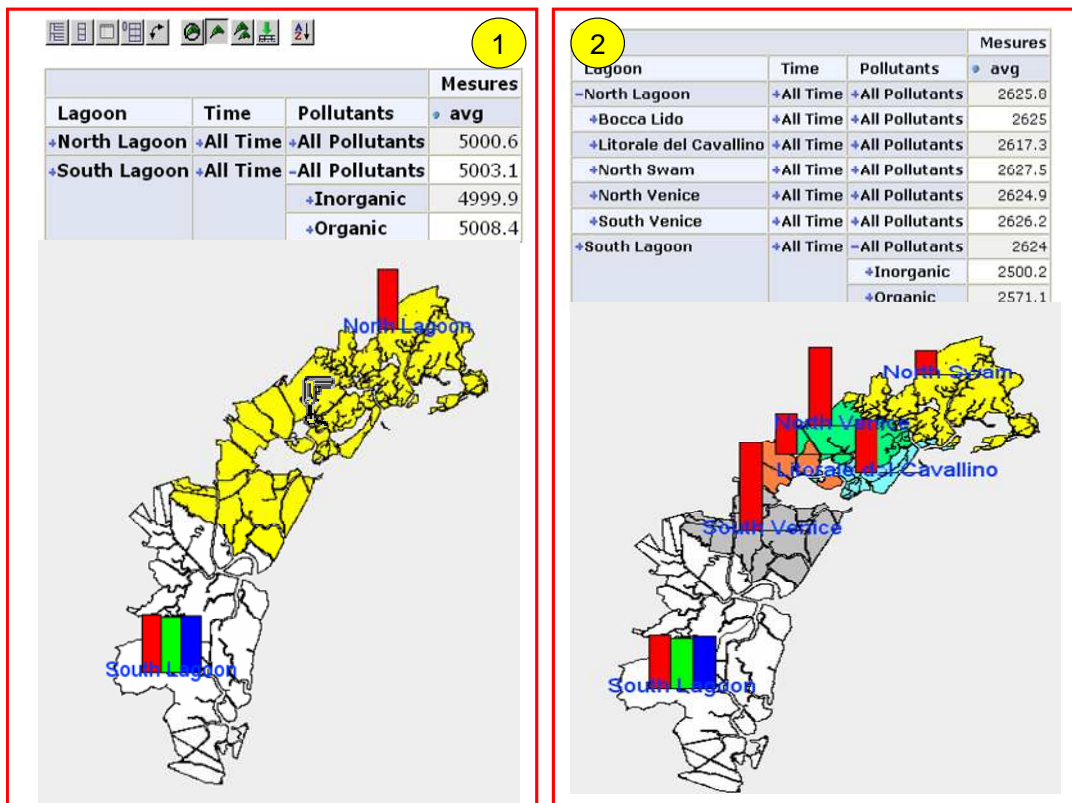


Figure 6.17. GeWolap drill-down position.

### 6.3.2.2 Opérateurs de coupe

Les opérateurs de coupe *GeWolap slice predicate* et *GeWolap slice position* implémentent l'opérateur Slice de GeoCube. Ils permettent de sélectionner un sous-ensemble de membres de la dimension géographique. Ceux-ci peuvent être spécifiés par pointage de la souris ou grâce à une requête qui utilise des prédicats spatiaux et/ou alphanumériques, à la différence de l'approche fournie par le *Cube Navigator* de JPivot où les membres doivent être sélectionnés parmi la liste de tous les membres des dimensions (Figure 6.12). Ces prédicats peuvent utiliser les seuls membres géographiques, aussi bien que toutes les données spatiales et/ou alphanumériques stockées dans la base de données.



Ces opérateurs sont nécessaires à l'analyse spatio-multidimensionnelle car ils représentent une traduction dans un contexte OLAP des méthodes d'analyse spatiale d'interrogation et de raisonnement présentées au chapitre 2. Supposons par exemple que l'utilisateur soit intéressé seulement par les zones de la lagune avec une surface supérieure à 1000 Km<sup>2</sup>. Alors, à partir de la configuration visuelle affichée en Figure 6.18-1, grâce au *GeWOLap slice predicate*, l'utilisateur formule une requête spatiale en utilisant l'extension SQL d'Oracle pour les données spatiales (Figure 6.19).

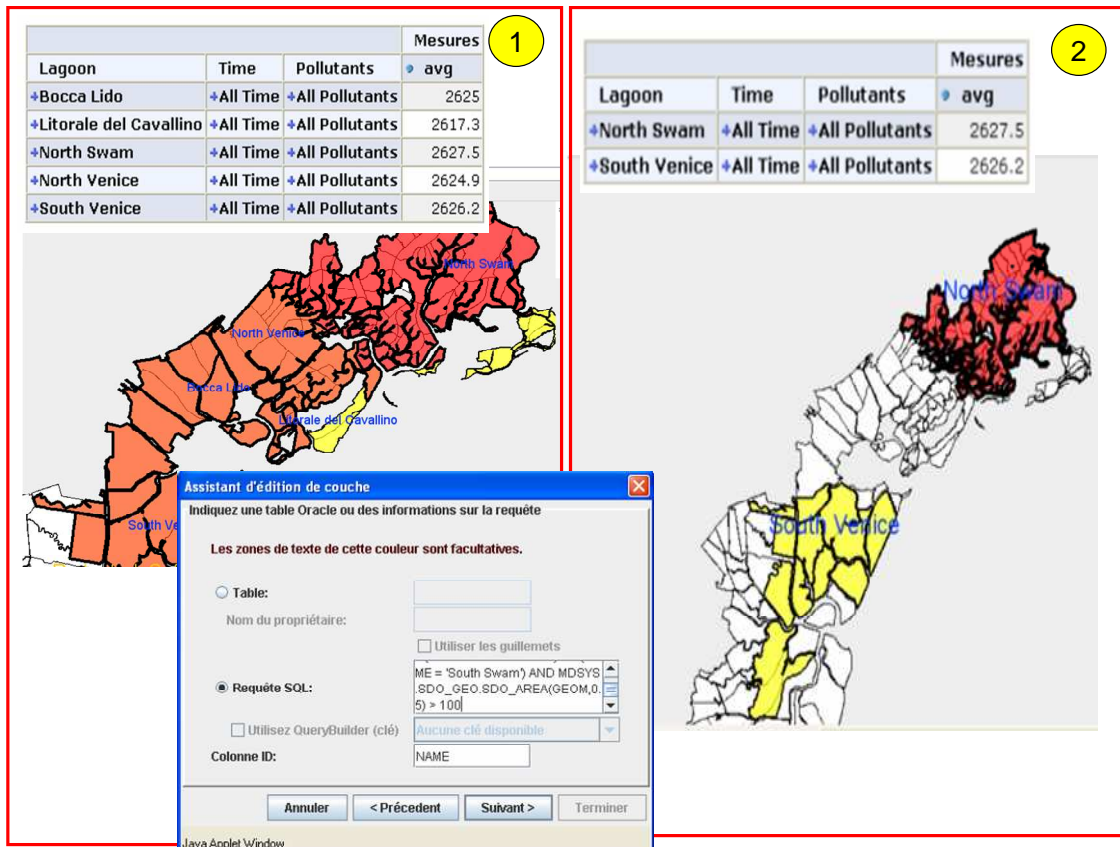


Figure 6.18. GeWOLap slice predicate : 1) Données initiales : toutes les régions de la lagune  
2) Résultat : régions de la lagune avec une surface supérieure à 1000 Km<sup>2</sup>.

Cette requête SQL est directement exécutée par Oracle. Ensuite, GeWOLap, en utilisant ce résultat, crée la requête MDX associée à l'opération de coupe, qui est enfin exécutée par Mondrian. Le résultat de l'opération de coupe est montré en Figure 6.18-2. Nous notons qu'avant l'opération de coupe, toutes les zones de la lagune étaient représentées dans la table de pivot et dans la carte. Après l'opération de coupe, seulement les zones avec une surface supérieure à 1000 Km<sup>2</sup>, c'est-à-dire « South Venice » et « North Swam », sont affichées à la fois dans la carte et dans la table de pivot.

```

select * from MEASURES_VIEW
where (NAME = 'Bocca Lido') OR(NAME = 'Bocca
Lido') OR(NAME = 'Bocca Lido') OR(NAME =
'Litorale del Cavallino') OR(NAME = 'Litorale
del Cavallino') OR(NAME = 'Litorale del
Cavallino') OR(NAME = 'North Swam') OR(NAME =
'North Swam') OR(NAME = 'North Swam') OR(NAME =
'North Venice') OR(NAME = 'North Venice')
OR(NAME = 'North Venice') OR(NAME = 'South
Venice') OR(NAME = 'South Venice') OR(NAME =
'South Venice') OR(NAME = 'South Lagoon')
OR(NAME = 'South Lagoon') OR(NAME = 'South
Lagoon') AND
MDSYS.SDO_GEOM.SDO_AREA(GEOM, 0.005) > 1000

```

Figure 6.19. Requête SQL représentant le prédicat spatial d'une opération de coupe.

En conclusion, les opérateurs *GeWolap slice predicate* et *GeWolap slice position* implémentent l'opérateur *Slice* de *GeoCube*. Ils permettent de couper l'hypercube grâce à des prédicats spatiaux et alphanumériques, étendant les opérateurs de coupe définis par *Mondrian* et *JPivot*.

### 6.3.2.3 Opérateurs de changement dynamique de l'hypercube

Les opérateurs de modification dynamique de l'hypercube *GeWolap buffer* et *GeWolap overlay* créent de nouveaux membres de dimensions géographiques grâce aux opérateurs d'analyse spatiale de *buffer* et *overlay*.

Le *GeWolap buffer* implémente l'opérateur *OLAP-Buffer* de *GeoCube*. Il permet d'exprimer des requêtes sur un membre de la dimension géographique défini à la volée par l'utilisateur, grâce à la désignation d'une zone tampon. Pour ce nouveau membre, la mesure doit être recalculée grâce à une fonction fournie par l'utilisateur. Le calcul de la mesure prend en compte les parties des membres recouverts totalement et/ou partialement par la région tampon, les membres originaux et les valeurs de mesures dans la table de faits associés. Supposons ainsi que la concentration des polluants soit calculée comme une moyenne pondérée sur les surfaces. La Figure 6.20-2 montre la pollution moyenne pour « Bocca Chioggia », « Chioggia », « Litorale Pellestrino », « South Swam » et une zone tampon de 3 km autour de « Bocca Malmocco ».

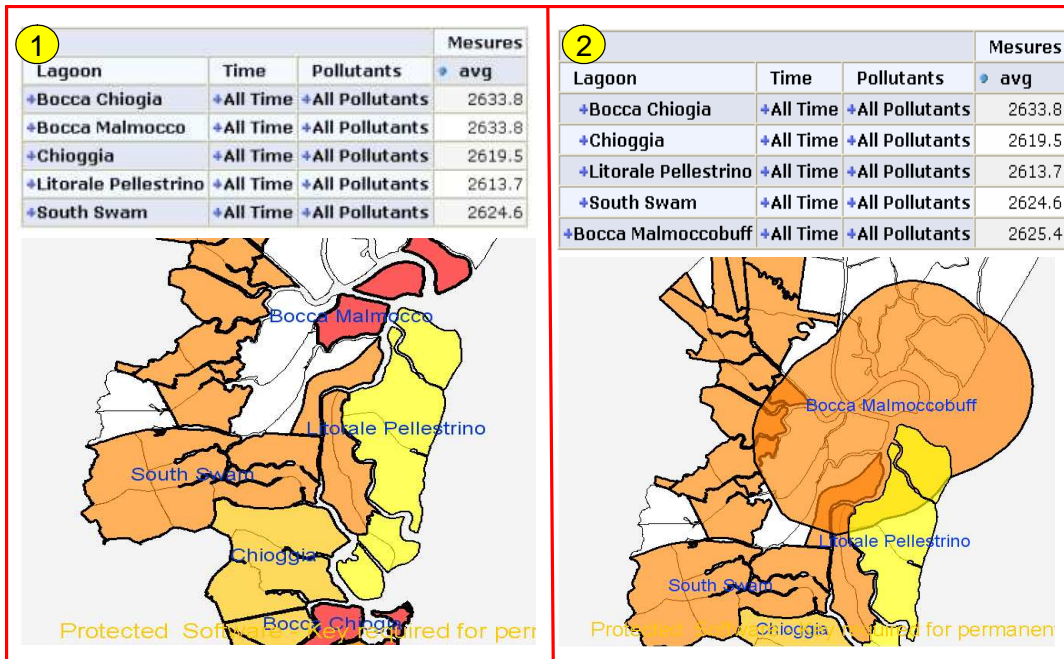


Figure 6.20. GeWolap buffer 1) Données initiales 2) Résultat.

L'opérateur *GeWolap overlay* permet de formuler des requêtes sur des membres calculés en utilisant une autre information géographique issue d'une autre couche et une fonction qui est fournie par l'utilisateur. Le calcul de la mesure prend en compte les parties de membres issus de l'opération d'overlay, les membres originaux des deux couches et les valeurs de mesures dans la table de faits associée. Supposons que l'utilisateur veuille ajouter aux informations de la pollution une autre source de données, par exemple une carte des zones de la lagune définissant les zones soumises à épuration. Présignons que la valeur de la nouvelle mesure puisse être estimée comme une moyenne pondérée sur la surface des zones non épurées. La Figure 6.21-2 montre les valeurs modifiées de la pollution pour les zones de la lagune « North Venice » soumises à épuration. Nous notons que la pollution dans « Isola delle Tresse » avant l'overlay était de 2631.4 (Figure 6.21-1). Le *GeWolap overlay* nous permet de mettre en évidence une zone d'épuration (« Isola delle Tresse-ZoneRD ») où la valeur de la pollution a pu être ré-estimée.

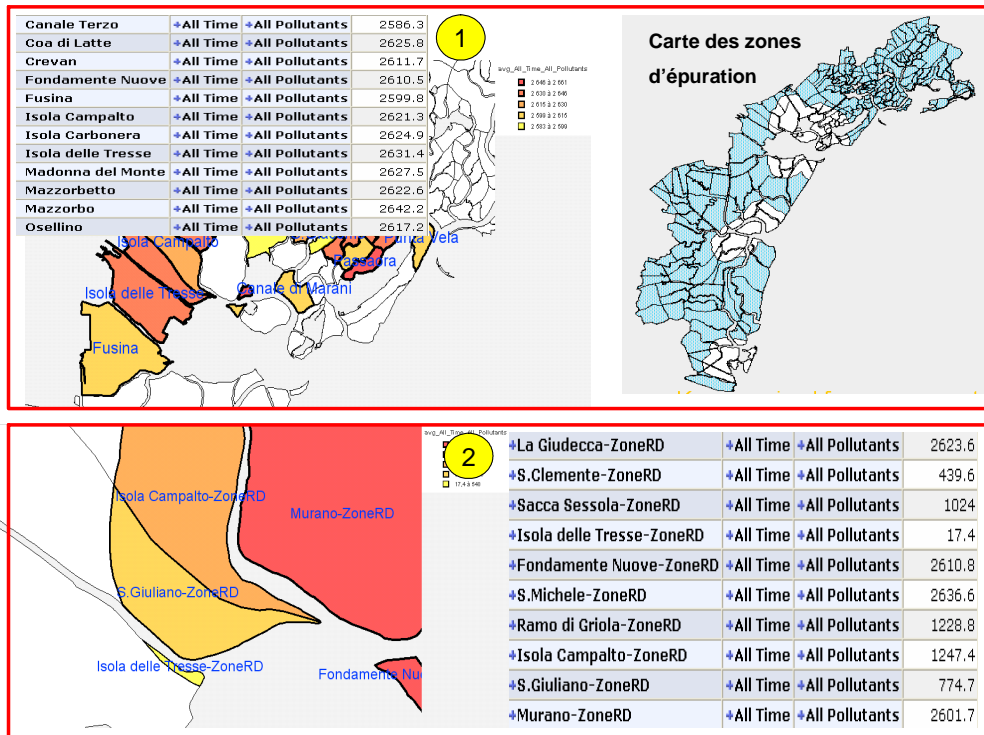


Figure 6.21. 1) Données initiales 2) Résultats du GeWolap overlay.

Pour implémenter ces opérateurs, GeWolap exploite la possibilité de créer et modifier le fichier XML représentant le schéma de l'application multidimensionnelle. L'idée est de créer pour chaque application multidimensionnelle une fonction permettant de recalculer les valeurs de la mesure. Il s'agit d'une méthode Java qui crée un nouveau schéma où les membres issus des opérations spatiales sont modélisés en tant que membres calculés. Une formule MDX, qui peut utiliser des fonctions java, est associée à ces membres calculés afin de recalculer les valeurs de la mesure. Une partie du fichier XML créé d'après l'opération de *GeWolap overlay* pour l'exemple précédent est montré en Figure 6.22.

```

...
<Measure name="avg" column="VALUE_READ" aggregator="avg"
  formatString="#.#" />

<CalculatedMember name="Isola delle Tresse-ZoneRD" dimension="Lagoon">
  <Formula>([Lagoon].[North Lagoon].[North Venice].[Isola delle
  Tresse]* 25170)/( 25170 + 0.7 *(5418719.641-25170))</Formula>
</CalculatedMember>
<CalculatedMember name="Fondamente Nuove-ZoneRD"
  dimension="Lagoon">
  <Formula>([Lagoon].[North Lagoon].[North Venice].[Fondamente
  Nuove]* 161000)/( 161000 + 0.7 *(160979.7073-161000))</Formula>
</CalculatedMember>
...

```

Figure 6.22. Schéma issu de l'opération de GeWOlap overlay.

### 6.4 Mesures complexes et géographiques, et opérateurs multidimensionnels

Dans cette section, nous nous intéressons aux mesures complexes et géographiques comme nous les avons définies dans GeoCube. Nous introduisons la modélisation logique utilisée pour la prise en compte de ces mesures dans une vision symétrique des mesures et des dimensions. Ensuite, nous montrons comment l'opérateur Roll-Up de GeoCube qui agrège les mesures complexes et géographiques, est implémenté dans GeWOlap.

Tout au long de cette section, nous reprenons les exemples concernant la pollution de la lagune de Venise utilisés au chapitre 5. Ces exemples permettent d'analyser les unités (mesures géographiques) en fonctions du temps, des polluants et de la valeur de pollution (Figure 6.23), et les polluants (mesures complexes) en fonction du temps, des unités (dimension géographique) et de la valeur de la pollution (Figure 6.24).

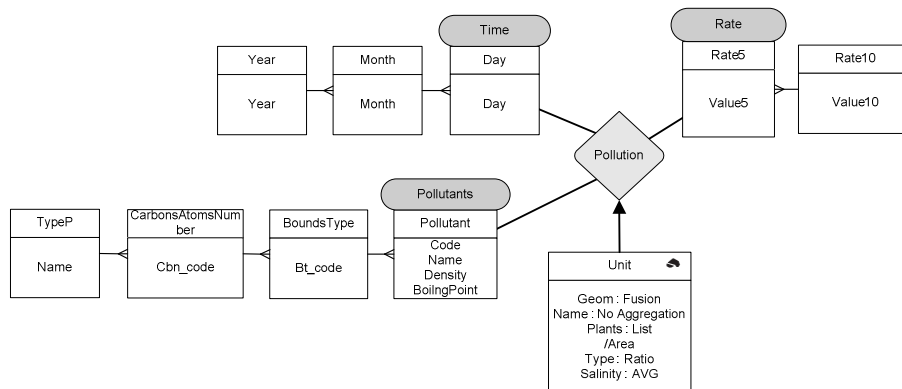


Figure 6.23. Application multidimensionnelle avec la mesure géographique « Unit ».

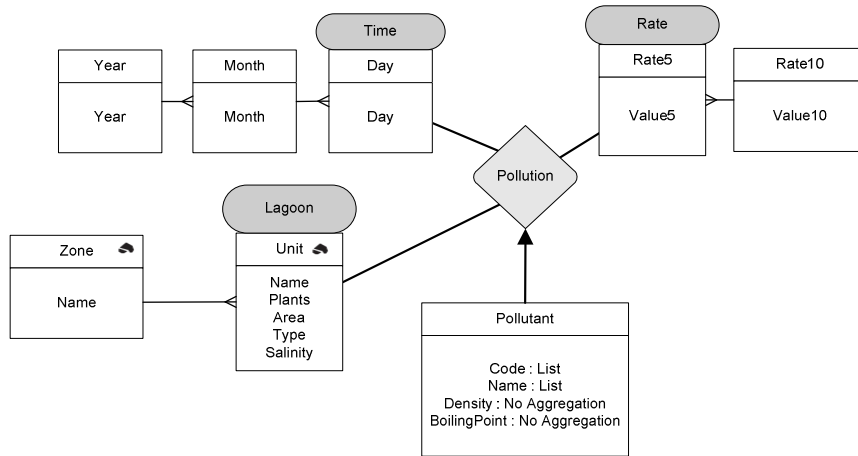


Figure 6.24. Application multidimensionnelle avec la mesure complexe « Pollutant » et la dimension géographique « Lagoon ».

### 6.4.1 Modélisation logique

GeWOlap utilise une table de faits qui ne fait pas de différence entre mesures et dimensions. Elle contient seulement les identifiants des objets complexes et/ou géographiques qui représentent les niveaux les plus détaillés de toutes les dimensions. En d'autres termes, il s'agit d'une table de faits sans faits. La symétrie entre mesure et dimension est éliminée seulement dans le schéma utilisé par Mondrian. Pour chaque mesure, un schéma différent est défini. Ces schémas présentent tous la même table de faits et les mêmes dimensions. Le squelette commun aux applications multidimensionnelles des Figures 6.23 et 6.24 est montré en Figure 6.25.

```

<xml version="1.0" ?>
<Schema name="CORILACUBE">
  <Cube name="POLLUTION_UNIT">
    <Table name="CORILAFACTS">
      Aggregate Table
    </Table>
    <Dimension name="Lagoon" foreignKey="UNIT_CODE">
      <Hierarchy hasAll="true" primaryKey="UNIT_CODE">
        <Table name="LOCATIONNDIM" />
        <Level name="Zone" column="ZONE" uniqueMembers="false" />
        <Level name="Unit" column="UNIT" uniqueMembers="false" />
      </Hiérarchie>
    </Dimension>
    <Dimension name="Rate" foreignKey="RATE5_CODE">
      <Hierarchy hasAll="true" allMemberName="All rate"
        primaryKey="RATE5_CODE">
        <Table name="RATEDIM" />
        <Level name="Rate10" column="RATE10" uniqueMembers="false" />
        <Level name="Rate5" column="RATE5" uniqueMembers="false" />
      </Hiérarchie>
    </Dimension>
    <Dimension name="Time" foreignKey="DAY_CODE">
      <Hierarchy hasAll="true" allMemberName="All Time"
        primaryKey="DAY_CODE">
        <Table name="TIMEDIM" />
        <Level name="Year" column="YEAR" uniqueMembers="false"/>
        ...
        <Level name="Day" column="DAY" uniqueMembers="false" />
      </Hiérarchie>
    </Dimension>
    <Dimension name="Pollutant" foreignKey="POLLUTANT_CODE">
      <Hierarchy hasAll="true" allMemberName="All Pollutant"
        primaryKey="POLLUTANT_CODE">
        <Table name="POLLUTANTDIM" />
        <Level name="Type" column="TYPE" uniqueMembers="false" />
        ...
        <Level name="Pollutant" column="POLLUTANT"
          uniqueMembers="false" />
      </Hierarchy >
    </Dimension>
    Mesure
  </Cube>
</Schema>

```

Figure 6.25. Squelette du schéma utilisé pour la modélisation des applications des figures 6.23 et 6.24.

Nous notons la présence de quatre dimensions « Time », « Lagoon », « Rate » et « Pollutant » et de la table de faits « CORILAFACTS », indépendamment de la mesure. Dans toutes les applications seulement la définition de la mesure dans la balise « mesure »

et la table agrégée changent. Les balises utilisées pour la mesure « Unit » et « Pollutant » sont montrées respectivement en Figure 6.26a et Figure 6.26b.

```
<Measure name="UNIT" column="UNIT_CODE"
aggregator="UserDefinedAggregation" />
```

(a)

```
<Measure name="POLLUTANT" column="POLLUTANT_CODE"
aggregator="UserDefinedAggregation" />
```

(b)

Figure 6.26. Balise Measure a) Mesure géographique b) Mesure complexe.

Même si les mesures sont différentes, la fonction d’agrégation est toujours la même « UserDefinedAggregation » (Figure 6.26). Cette fonction appelle un mécanisme de mapping, implémenté dans Oracle à travers une fonction d’agrégation ad hoc définie en PL/SQL, entre les identifiants des mesures détaillées et l’identifiant de la mesure agrégée. Dans notre solution, Mondrian délègue totalement à Oracle le processus d’agrégation. Le lien entre Mondrian et Oracle est illustré en Figure 6.27.

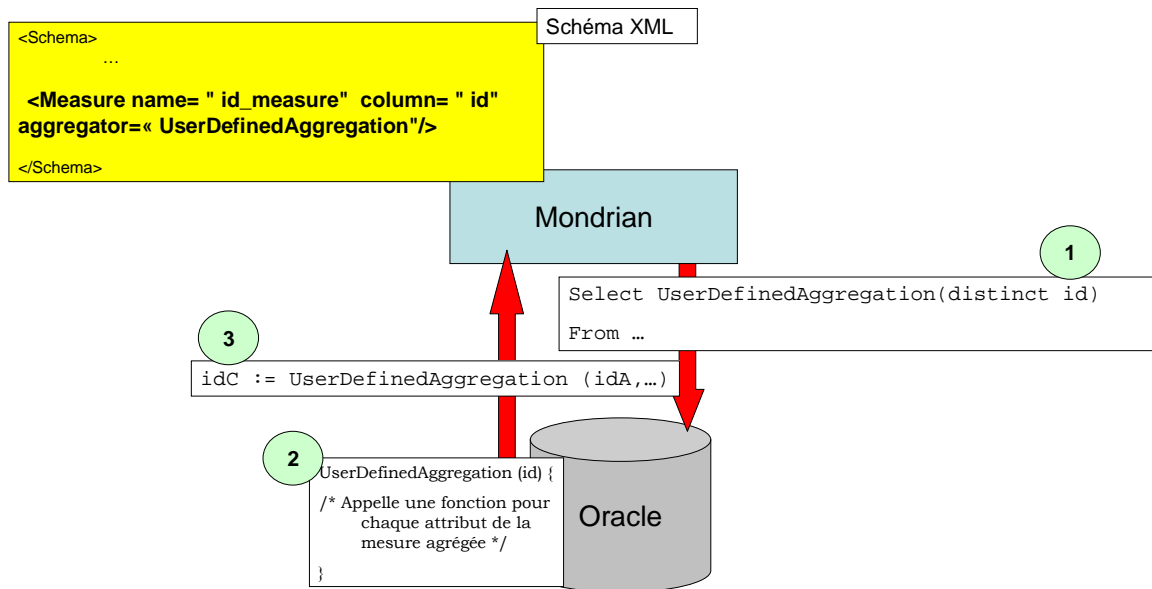


Figure 6.27. Processus d’agrégation des mesures complexes et géographiques.

Grâce au fichier XML qui contient la fonction d’agrégation « UserDefinedAggregation », définie sur l’attribut représentant l’identifiant de la mesure, Mondrian traduit toutes les requêtes multidimensionnelles en requêtes SQL qui utilisent cette fonction d’agrégation à la place des fonctions SQL classiques (Figure 6.27-1). Cette



requête est prise en charge par Oracle, qui déclenche l'agrégation des attributs des mesures détaillées, dont les identifiants doivent être agrégés par `UserDefinedAggregation`, créant une nouvelle mesure agrégée (Figure 6.27-2). L'identifiant de cette mesure agrégée est le résultat de l'agrégation `UserDefinedAggregation` (Figure 6.27-3). Le résultat est ensuite renvoyé au client OLAP.

Comme nous l'avons montré aux chapitres 4 et 5, la prise en compte des mesures multi-valuées est une caractéristique importante de l'OLAP Géographique. Cet aspect n'est pas modélisé explicitement dans la table de faits sans faits utilisée par GeWOLap, car chaque cellule de la table de faits contient l'identifiant d'un seul membre de dimension. Pour remédier à cette limitation, nous utilisons le mécanisme des vues matérialisées fourni par Mondrian. Pour chaque mesure, nous définissons une table agrégée qui enlève la dimension qui contient la mesure, et qui maintient toutes les autres dimensions aux niveaux les plus détaillés. Par conséquent, les requêtes qui portent sur le cuboïde qui utilise les niveaux définis dans la table agrégée, n'agrègent pas les mesures avec la fonction d'agrégation « `UserDefinedAggregation` », mais se limitent à utiliser les valeurs des mesures présentes dans la table agrégée. Celles-ci, à la différence des mesures dans la table des faits, contiennent un pointeur vers une liste d'objets complexes ou géographiques. Cette modélisation logique permet à GeWOLap de gérer les mesures multi-valuées. Les tables agrégées utilisées pour les mesures « Unit » et « Pollutant » sont décrites respectivement en Figure 6.28a et Figure 6.28b

```

<AggName name="AGGFACTSUNIT">
  <AggFactCount column="FACT_COUNT" />
  <AggMeasure name="[Measures].[UNIT]" column="UNIT_CODE" />
  <AggLevel name="[Pollutant].[Type]" column="TYPE_CODE" />
  ...
  <AggLevel name="[Pollutant].[Pollutant]" column="POLLUTANT_CODE" />
  <AggLevel name="[Rate].[Rate10]" column="RATE10_CODE" />
  <AggLevel name="[Rate].[Rate5]" column="RATE5_CODE" />
  <AggLevel name="[Time].[Year]" column="YEAR_CODE" />
  <AggLevel name="[Time].[Month]" column="MONHT_CODE" />
  <AggLevel name="[Time].[Day]" column="DAY_CODE" />
</AggName>

```

(a)

```

<AggName name="AGGFACTSPOLLUTANT">
  <AggFactCount column="FACT_COUNT" />
  <AggMeasure
name="[Measures].[POLLUTANT]"column="POLLUTANT_CODE" />
  <AggLevel name="[Location].[Zone]" column="ZONE_CODE" />
  <AggLevel name="[Location].[Unit]" column="UNIT_CODE" />
  <AggLevel name="[Rate].[Rate10]" column="RATE10_CODE" />
  <AggLevel name="[Rate].[Rate5]" column="RATE5_CODE"
/><AggLevel name="[Time].[Year]" column="YEAR_CODE" />
  <AggLevel name="[Time].[Month]" column="MONHT_CODE" />
  <AggLevel name="[Time].[Day]" column="DAY_CODE" />
</AggName>

```

(b)

Figure 6.28. Aggregate Tables pour la gestion de mesures multi-valuées a) « Unit » b) « Pollutant ».

De plus, grâce à cette modélisation logique particulière et au mécanisme d'agrégation, le changement d'une mesure avec une dimension, se réduit au simple changement du fichier XML.

## 6.4.2 Visualisation et Navigation

Dans la section 6.2, nous avons décrit comment GeWOlap utilise les opérateurs de forage de JPivot pour agréger les mesures numériques. Dans les deux paragraphes suivants, nous illustrons l'implémentation de l'opérateur Roll-Up de GeoCube dans le cas des mesures complexes et géographiques, en nous concentrant sur les aspects de visualisation.

### 6.4.2.1 Mesures géographiques

L'introduction de l'information géographique en tant que sujet de l'analyse implique une reformulation des critères de visualisation des résultats des requêtes multidimensionnelles. Les valeurs des cellules de la table de pivot ne contiennent plus des valeurs numériques, mais une carte représentant l'ensemble des mesures détaillées ou une mesure agrégée. Une correcte visualisation des mesures géographique soulève des nombreuses problématiques et, comme nous le décrivons au chapitre suivant, des nouvelles techniques de visualisation et d'interaction sont nécessaires. Dans notre prototype nous adoptons une solution simple où chaque cellule de la table de faits contient les identifiants

des mesures géographiques (détaillées ou agrégées). Ces descriptions textuelles des objets géographiques sont insuffisantes pour une analyse spatio-multidimensionnelle satisfaisante. La composante cartographique de l'interface alors est utilisée pour visualiser les mesures géographiques. Elle présente une carte, représentant l'ensemble de toutes les mesures et, contrairement aux dimensions géographiques, et une autre carte dont les éléments sont les mesures géographiques choisies dynamiquement par l'utilisateur. La table de pivot qui représente le cuboïde de base de l'application de la Figure 6.23 est affichée en Figure 6.29. Elle permet de répondre à la requête : « *Quelles sont les unités polluées pour chaque jour, pour chaque polluant, et pour chaque valeur de pollution ?* ».

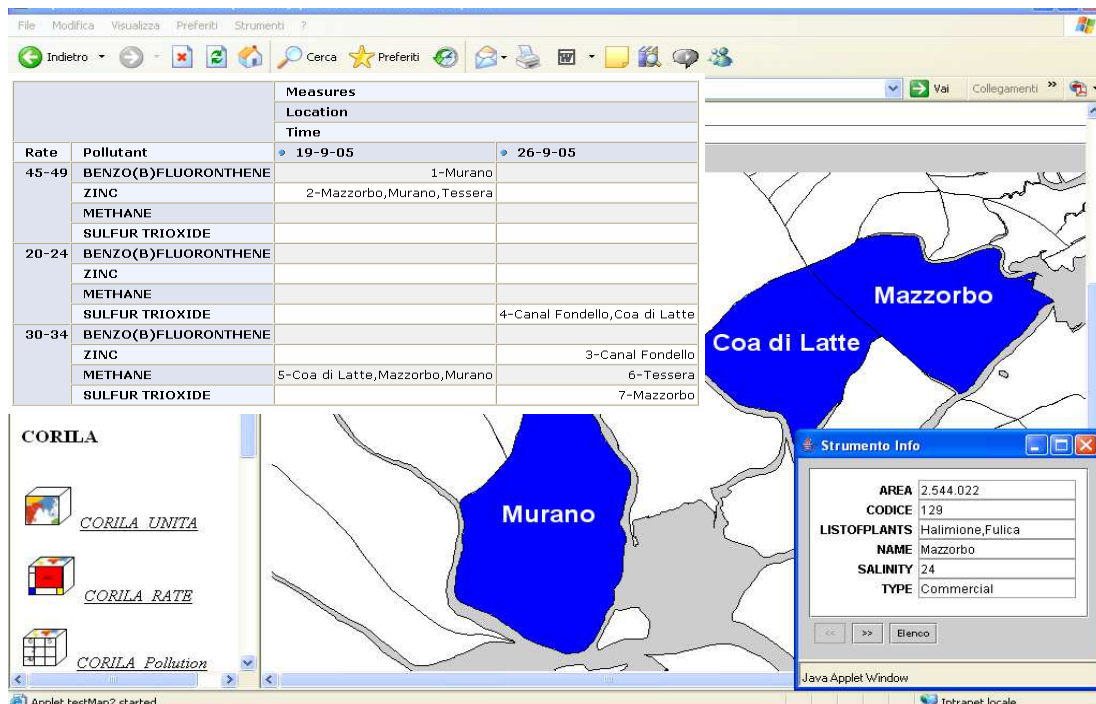


Figure 6.29. Mesure géographique « Unit ».

La table de pivot contient des valeurs numériques, qui représentent des ensembles d'unités. Par exemple le 19-9-05, le méthane était présent à une concentration de 30-34 mg/l dans l'ensemble d'unités avec identifiant 5. L'utilisateur sélectionne la cellule (i.e. la carte) et GeWOLap montre alors la carte représentant toutes les unités de la lagune en blanc, et en bleu les unités polluées : « Murano », « Coa di Latte » et « Mazzorbo ». Une mesure géographique est caractérisée par un ensemble d'attributs alphanumériques. En cliquant sur « Mazzorbo » par exemple, l'utilisateur récupère toutes ses valeurs alphanumériques : la surface, la liste de plantes, le nom, la salinité et le type d'unité, ceci grâce à l'*Info Tool* dans la toolbar SIG. Supposons que l'utilisateur décide d'agréger les unités en montant au niveau mois de la dimension temporelle. Grâce au *Cube Navigator*, il sélectionne le mois de septembre de 2005. Cette action correspond à une opération de roll-up comme défini par l'opérateur Roll-up de GeoCube. Elle déclenche l'agrégation des mesures géographiques à travers la fonction d'agrégation UserDefinedAggregation, et un ensemble de fonctions d'agrégation ad hoc pour agréger les attributs spatiaux et alphanumériques des unités : l'union topologique pour la géométrie, la moyenne pour la salinité, une liste pour les

plantes et une fonction particulière pour le type de zone. Le résultat du Roll-Up est affiché en Figure 6.30. La table de pivot ne contient plus les jours mais le membre « September 2005 ». Les valeurs de la table de pivot sont les identifiants des mesures géographiques agrégées. Par exemple, le 9-05, la zone agrégée avec identifiant « 231418 » est polluée au méthane avec une valeur de 30-34 mg/l. La géométrie de cette mesure géographique est l'union topologique des géométries des mesures détaillées : « Coa di Latte », « Mazzorbo », « Murano » et « Tessera ». Nous notons, grâce à l'Info Tool, que cet objet géographique n'a pas de nom, que la salinité, calculée en utilisant une moyenne, est de 25.5, la liste des plantes contient toutes les plantes de toutes les unités, et que le type est « commercial ».

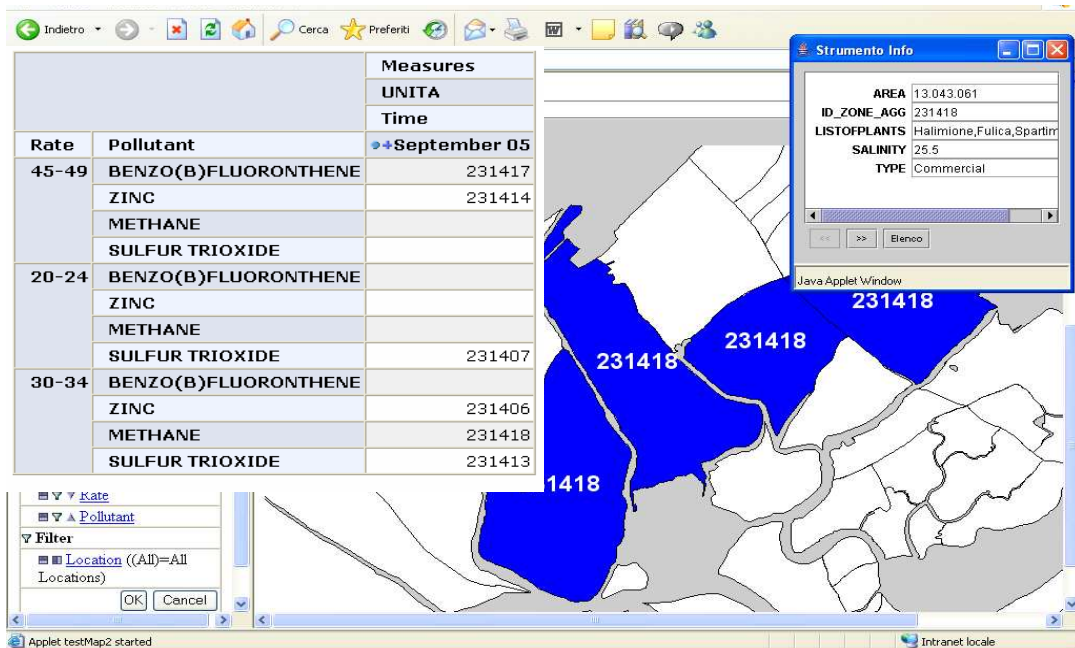


Figure 6.30. Agrégation des mesures géographiques « Unit ».

GeWolap permet ainsi de visualiser et d'agréger les mesures géographiques, selon l'approche formelle définie par l'opérateur Roll-Up de GeoCube, en étendant les opérateurs de forage fournis par JPivot-Mondrian.

#### 6.4.2.2 Mesures complexes

GeWolap, basé sur GeoCube, gère les mesures de type géographique mais aussi de type complexe. Considérons l'application multidimensionnelle de la Figure 6.24 avec les polluants comme mesures. Cette application présente une dimension géographique et une mesure complexe. La visualisation et l'interaction avec la dimension géographique restent identiques à ce que nous avons décrit en section 2. La visualisation des mesures dans la table de pivot est similaire à celle des mesures géographiques. Les identifiants des mesures sont affichés dans les cellules de la table de pivot. L'affichage des mesures sur les membres de la dimension géographique change. Les techniques de visualisation montrées en section 6.2 ne sont pas applicables car un objet complexe n'est pas une simple valeur numérique mais un ensemble indivisible de valeurs alphanumériques. Afin de visualiser les mesures complexes sur une dimension géographique, nous utilisons un ensemble de cartes

choroplèthes. Chaque objet ou ensemble d'objets géographiques est représenté avec une couleur, car cette variable visuelle est adaptée pour la représentation des valeurs nominales (Bertin et Bonin, 1992). Pour clarifier ce concept, nous présentons le résultat d'une navigation dans GeWOLap qui répond à la requête : « *Quels sont les polluants qui polluent les unités de la lagune, pour chaque jour et pour chaque valeur de pollution ?* ».

La table de pivot est montrée en Figure 6.31. Nous notons que les données multidimensionnelles de cette table sont les mêmes que celles de la table de pivot de figure 6.29.

		Measures	
		POLLUTANT	
		Time	
Location	Rate	• 19-9-05	• 26-9-05
Mazzorbo	45-49		1-ZINC
	20-24		
	30-34	2-SULFUR TRIOXIDE,METHANE	
Coa di Latte	45-49		
	20-24		3-SULFUR TRIOXIDE
	30-34	4-METHANE	
Tessera	45-49	1-ZINC	
	20-24		
	30-34		4-METHANE
Canale Fondello	45-49		
	20-24		3-SULFUR TRIOXIDE
	30-34	1-ZINC	
Murano	45-49	5-BENZO(B)FLOURONTHENE,ZINC	
	20-24		
	30-34	4-METHANE	

Figure 6.31. Table de pivot pour la mesure complexe « Pollutant ».

Pour cette table de pivot, GeWOLap génère une carte choroplèthe pour chaque combinaison des membres des dimensions non géographiques : (45-49 ; 19-9-05), (45-49 ; 26-9-05) ; (30-34 ; 19-9-05), (30-34 ; 26-9-05), (20-24 ; 19-9-05), (20-24 ; 26-9-05). La carte thématique pour la combinaison des membres « 30-34 » mg/l et « 19-9-05 » est montrée en Figure 6.32. Cette carte thématique montre que les unités « Coa di Latte » et « Murano » sont polluées par les mêmes polluants, et que « Tessera » n'est pas polluée. L'utilisateur sélectionne par exemple les polluants pour l'unité Mazzorbo. GeWOLap alors représente, à travers une table, la liste des polluants avec tous leurs attributs alphanumériques : le code, le nom, la densité et le point d'ébullition.

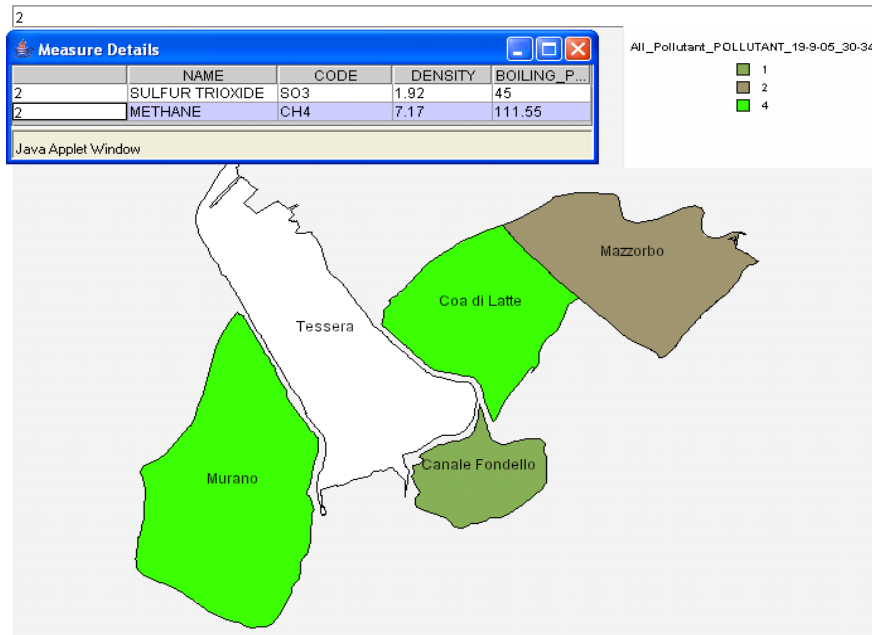


Figure 6.32. Carte choroplèthe pour la mesure complexe « Polluant » associée au jour 19/9/05 et à la valeur de pollution 30-34 mg/l.

La carte thématique pour la combinaison des membres « 30-34 » mg/l et « 26-9-05 » est affiché en Figure 6.33.

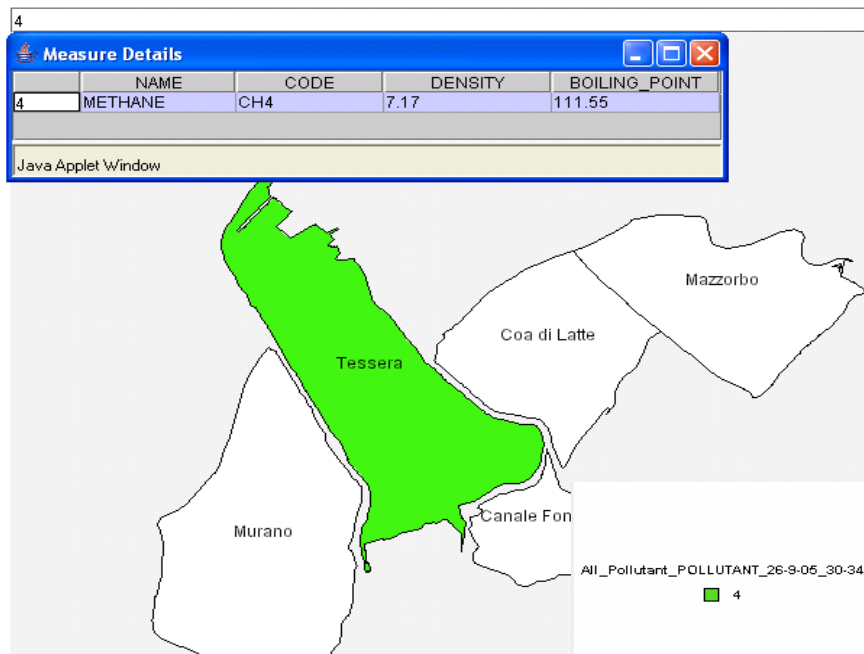


Figure 6.33. Carte choroplèthe pour la mesure complexe « Polluant » pour le jour 26/9/05 et la valeur de pollution 30-34 mg/l.

Cette représentation cartographique nous permet de visualiser l'évolution temporelle de la pollution dans la lagune et de voir qu'à distance d'une semaine la pollution change complètement dans l'unité Tessera.

## 6.5 Opérateurs SIG

Les opérateurs SIG intègrent et complètent les fonctionnalités multidimensionnelles d'un système d'analyse spatio-multidimensionnelle. GeWOLap met à disposition des fonctionnalités SIG de base comme le *Zoom (in et out)*, le déplacement de la carte, l'*Info Tool* et le *Control Layer*. Le *Control Layer* permet de personnaliser l'affichage des cartes et surtout d'ajouter des couches dans la composante cartographique de GeWOLap. Cette fonctionnalité, comme cela a été montré dans plusieurs travaux (Bédard, et al., 2001), (Matias et Moura-Pires, 2005) est indispensable car elle permet de visualiser le contexte géographique de l'analyse multidimensionnelle. En effet, lors de la conception du cube, seulement les données géographiques qui représentent les membres d'un axe d'analyse sont stockées dans l'entrepôt de données. En revanche, le processus d'analyse spatiale est un processus itératif et les données dont le décideur a besoin ne peuvent pas être fixées a priori. L'utilisateur peut alors avoir besoin de superposer aux membres géographiques d'autres informations géographiques pour valider ou reformuler visuellement ses hypothèses. De plus, grâce au *Control Layer*, il a accès à la base de données spatiales, pour analyser les données à travers des techniques d'analyse spatiale d'interrogation et de raisonnement. En d'autres termes, il peut interroger la base de données spatiales à travers de requêtes SQL.

## 6.6 Bilan et comparaison

Dans cette section, nous présentons les caractéristiques principales de GeWOLap, en comparant ce dernier aux solutions OLAP-SIG intégrées, présentées au chapitre 3. Enfin, nous mettons en évidence les limites de notre prototype.

### 6.6.1 Caractéristiques principales

GeWOLap est une solution web OLAP-SIG intégrée basée sur un SGBD spatial objet-relationnel, un serveur OLAP et une interface qui unifie un client OLAP et un client SIG. GeWOLap, grâce à Mondrian, fournit toutes les fonctionnalités avancées d'un serveur OLAP, comme les hiérarchies complexes, les propriétés des membres des dimensions, les hiérarchies multiples, les mesures dérivées, les vues matérialisées, etc. De plus, il met à disposition de l'utilisateur, parallèlement aux opérateurs multidimensionnels, des fonctionnalités typiquement SIG, comme l'ajout de couches, la personnalisation de leur affichage, un outil pour calculer la distance entre deux points, etc.

GeWOLap est basé sur le modèle formel GeoCube. Il fournit une vision symétrique des mesures et des dimensions, permettant de gérer, outre les mesures classiques numériques, les mesures complexes et géographiques et les dimensions géographiques. Les opérations de navigation multidimensionnelle de forage et de coupe implémentent les opérateurs Roll-Up et Slice de GeoCube. De plus, GeWOLap introduit deux nouveaux opérateurs qui adaptent au contexte multidimensionnel les opérateurs d'analyse spatiale d'overlay et de buffer.

L'interface web de GeWOLap est conçue spécialement pour supporter l'analyse spatio-multidimensionnelle dans le cas où l'hypercube contient des dimensions géographiques et des mesures numériques ou géographiques ou complexes. Elle combine et synchronise dans un seul environnement interactif trois composantes différentes : cartographique, tabulaire et graphique. Chaque interaction avec la composante tabulaire ou cartographique est traduite en une opération de navigation multidimensionnelle ou SIG. En particulier, la composante cartographique est une carte interactive qui adapte son affichage au type et au nombre de mesures.

### 6.6.2 Comparaison et limites

GeWOLap, basé sur le modèle GeoCube, présente deux avantages fondamentaux par rapport aux solutions SOLAP présentées au chapitre 3 : le support (visualisation et opérateurs multidimensionnels) pour les mesures géographiques et complexes dans une vision symétrique des mesures et des dimensions, et les opérateurs qui modifient dynamiquement la structure de l'hypercube en utilisant les opérateurs d'analyse spatiale. En effet, (Rivest et al., 2005), (JMap, 2007), (Scotch, et Parmanto, 2006) et (Matias et Moura-Pires, 2007) gèrent seulement les mesures comme objets spatiaux. Les opérateurs SIG définis dans (Rivest et al., 2005) et (Matias et Moura-Pires, 2007) sont des opérateurs orthogonaux aux opérateurs OLAP, et l'opérateur de buffer présenté dans (Scotch, et Parmanto, 2006) est utilisé seulement pour couper l'hypercube.

GeWOLap présente différentes limites liées à l'implémentation des opérateurs de navigation définis par GeoCube, et à l'implémentation d'autres fonctionnalités de visualisation.

Plus précisément, l'opérateur Dice, réalisable à travers une simple opération de projection, doit être encore développé. L'opérateur Slice de GeWOLap étend l'opérateur de coupe fourni par JPivot-Mondrian, car il permet d'utiliser des prédicats spatiaux et alphanumériques pour sélectionner les membres qu'on souhaite visualiser. En revanche, la propagation de l'opération de coupe à la table des faits, comme définie par GeoCube, doit être implémentée. Par exemple, suite à une opération de coupe sur le zinc, si l'utilisateur applique une opération de Roll-up sur tous les polluants, alors les mesures associées à tous les polluants seront prises encore en compte dans le calcul de la mesure agrégée, contrairement à l'approche définie par GeoCube qui utilisera seulement les mesures associées au zinc. De plus, dans GeWOLap l'opérateur de Slice peut être appliqué seulement aux membres des dimensions. L'extension de cet opérateur aux mesures nous permettra aussi d'implémenter les opérateurs Specialize et Classify comme des opérateurs de navigation entre différents hypercubes définis a priori.

Du point de vue visualisation, GeWOLap ne permet pas de gérer le « Multiple Windows Coordination » (cf chap 3 § 4.1.3), qui est une caractéristique importante des systèmes d'exploration et d'analyse des données. En effet, dans GeWOLap les trois composantes de l'interface visuelle sont accessibles uniquement à travers une seule fenêtre.

Enfin, un cadre global qui décide automatiquement les variables visuelles à utiliser, selon des règles sémiologiques, et la définition de nouveaux paradigmes de visualisation et



d'interaction pour l'analyse des données spatio-multidimensionnelles représentent des perspectives de recherche intéressantes.

## **6.7 Conclusions**

Dans ce chapitre, nous avons présenté GeWolap une solution web OLAP-SIG intégrée. GeWolap est basé sur le modèle GeoCube. GeWolap gère les mesures géographiques et complexes, et les dimensions géographiques. Il implémente un ensemble d'opérateurs de l'algèbre de GeoCube. GeWolap présente toutes les caractéristiques fondamentales de visualisation et d'interaction pour une analyse spatio-multidimensionnelle effective, même si la prise en compte des mesures géographiques nécessite une reformulation des techniques de visualisation et d'interaction existantes, comme cela est décrit au chapitre suivant.



# *Chapitre 7 : GeOlaPivot Table, une technique de visualisation et d'interaction pour l'analyse des mesures géographiques*

GeWOlap, comme nous avons décrit au chapitre 6, utilise une simple représentation textuelle pour représenter les mesures géographiques dans les cellules de la table de pivot. Cette approche n'est pas adéquate pour une analyse spatio-multidimensionnelle correcte, car comme nous l'avons détaillée au chapitre 4, les problématiques liées à la visualisation des mesures géographiques sont complexes et des techniques de visualisation et d'interaction particulières sont nécessaires.

Dans ce chapitre, nous présentons alors un nouveau paradigme de visualisation et d'interaction qui combine les concepts de table de pivot et Space-Time Cube : GeOlaPivot Table (Bimonte et al., 2006c) (Bimonte et al., 2007a). Dans la section 1, nous introduisons les concepts de Pivot Table et Space-Time Cube. Dans la section 2, nous présentons GeOlaPivot Table à travers une maquette de l'interface qui étend celle de JPivot. Nous concluons le chapitre avec une discussion sur les points forts et faibles de notre approche.

## **7.1 Pivot Table et Space-Time Cube**

Le paradigme le plus utilisé pour l'analyse des données multidimensionnelles est la table de pivot. (Stolte et al., 2002), (Tableau, 2007) couplent cette technique visuelle avec des affichages graphiques, ce qui permet d'enrichir le pouvoir expressif de la table de pivot grâce à d'importantes capacités de communication et de visualisation associées aux affichages graphiques. Ces solutions transforment les cellules de la table de pivot en un ensemble d'affichages graphiques. Chaque cellule montre alors les valeurs des mesures numériques associées à différentes combinaisons des membres de dimensions. Un exemple est montré en Figure 7.1, où dans chaque cellule, est affichée la somme des produits vendus pour chaque mois à travers des barres de différentes couleurs. Cette technique permet de comparer facilement les données multidimensionnelles, toujours en permettant à l'utilisateur d'explorer l'hypercube à travers le paradigme d'interaction de la table de pivot.

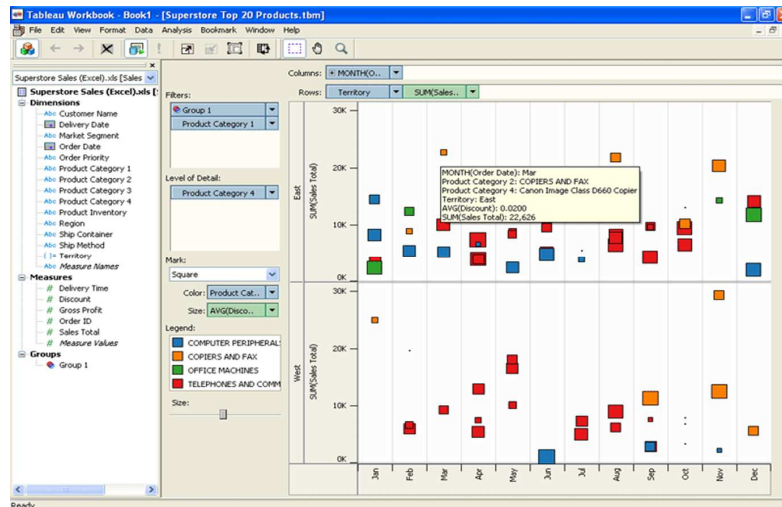


Figure 7.1 . Extension de la Table de Pivot avec des affichages graphiques et des variables visuelles dans Tableau Software.

La composante temporelle est souvent liée aux données spatiales. Pour étudier le rapport entre l'espace et le temps, Hägerstrand développe le concept de Space-Time Cube (Hägerstrand, 1970). Il s'agit d'une simple représentation graphique où la base du cube représente l'espace et l'axe vertical le temps. Les points dans le cube sont les données spatio-temporelles. Le Space-Time Cube permet de comprendre les trajectoires décrites par un phénomène naturel ou par une activité humaine. Un exemple de Space-Time Cube représentant le trajet d'une personne dans une journée est montré en Figure 7.2a. Les progrès accomplis dans le domaine de la géovisualisation ont donné une nouvelle vie à cette méthodologie. Le Space-Time Cube a été implémenté dans différents travaux (Gatalsky et al., 2004), (Kraak, 2003), (Dykes et Mountain, 2003) qui permettent de visualiser et d'explorer les données spatio-temporelles dans un contexte interactif pour la découverte des connaissances spatiales. (Gatalsky et al., 2004) utilisent cette technique de géovisualisation pour l'exploration des données des tremblements de terre (Figure 7.2b). Ils utilisent le Space-Time Cube non pas pour l'analyse des trajectoires, mais pour la découverte des tendances spatio-temporelles, comme des clusters d'événements. Dans un premier temps, les auteurs montrent que les techniques classiques de visualisation ne sont pas adaptées à l'analyse des données spatio-temporelles car elles peuvent cacher et/ou modifier les vraies données. Ensuite, ils montrent comment, grâce au Space-Time Cube, est obtenue une vision claire et montrant un cluster de tremblements de terres.

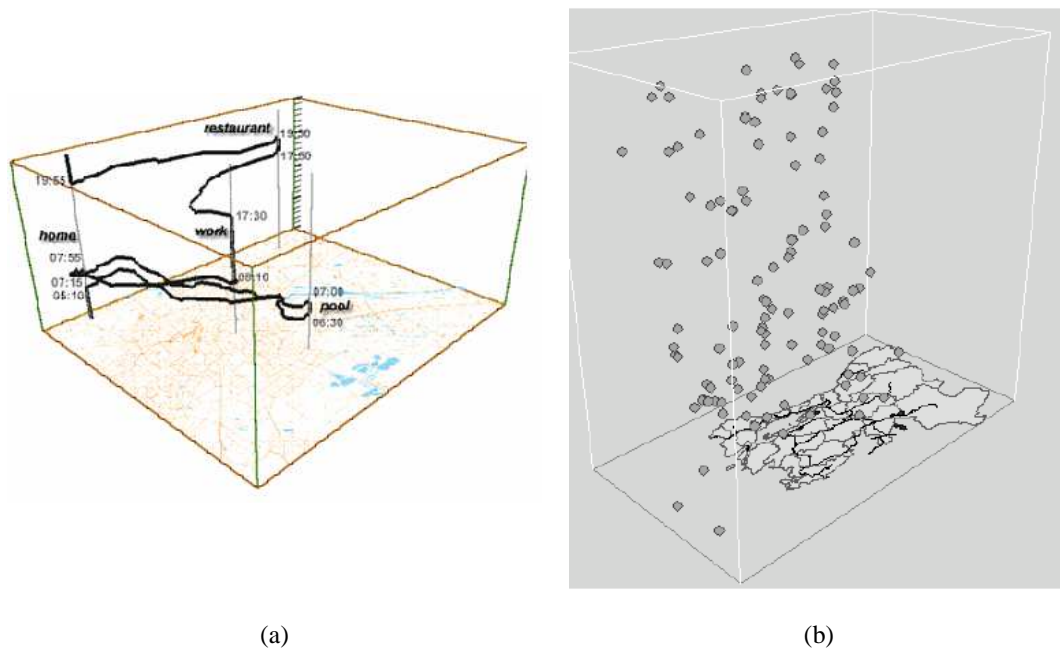


Figure 7.2. Space-Time Cube a) Etude des trajectoires (Kraak, 2003) b) Etude des tremblements de terre (Gatalsky et al., 2004).

## 7.2 GeOlaPivot Table

Dans cette section, nous décrivons GeOlaPivot Table, à l'aide d'une maquette d'interface visuelle.

### 7.2.1 Le concept

Nous proposons une technique de visualisation et d'interaction qui combine les concepts de table de pivot et de Space-Time Cube. Cette solution est inspirée de l'approche « Query by example » de Zloof (Zloof, 1975). L'idée de GeOlaPivot Table est d'inclure dans chaque cellule de la table de pivot un Space-Time Cube. L'axe vertical du cube représente un niveau d'une dimension, qui, contrairement à l'approche classique, peut ne pas être la dimension temporelle, à savoir, par exemple, le niveau représentant les polluants ou les produits. La base du cube est représentée par un membre géographique, s'il existe, et les points du cube sont les mesures géographiques. Les mesures associées à la même combinaison des membres de dimensions sont affichées avec la même couleur. De plus, les attributs alphanumériques sont visualisés en utilisant des affichages graphiques (i.e. barres, camemberts, etc.). Un cube peut être tourné pour obtenir le meilleur point de vue, évitant les encombrements visuels et informatifs. Un exemple de GeOlaPivot Table pour répondre à la question « *Quelles sont les provinces du nord d'Italie (« Northern Italy ») touchées par le SIDA pour chaque année (« Year ») avec un taux de mortalité supérieure à 6.0 (« Incidence ») ?* » est montrée en Figure 7.3.

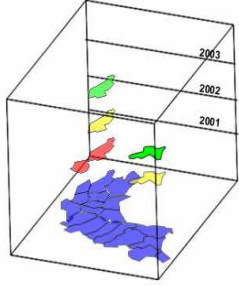
<i>Incidence</i>	<i>Location</i>	<i>Year</i>
> 6.0	+ Northern Italy	

Figure 7.3. Exemple de GeOlaPivot Table.

La Table de pivot présente une dimension numérique « Incidence » et une dimension géographique « Location ». La cellule associée à la combinaison des membres (« > 6.0 » ; « Northern Italy ») est un cube où la base est l'objet géographique « Northern Italy », les membres « 2001 », « 2002 » et « 2003 » du niveau « Year » sont présents sur l'axe vertical et les points du cube sont les mesures géographique de l'application multidimensionnelle : les provinces.

GeOlaPivot Table fournit une visualisation unique et interactive d'un ensemble de mesures géographiques associées à un ensemble de cellules de l'hypercube spatial, exploitant les trois dimensions du cube. Deux dimensions sont utilisées pour la représentation de la composante spatiale des données (mesures et membres de dimensions) et la troisième dimension permet de visualiser ces données en fonction d'une autre dimension de l'application multidimensionnelle. Le cube permet de visualiser le fait que la même mesure géographique est associée à différentes combinaisons de membres de dimensions grâce à l'axe vertical exclusivement sans utiliser de représentations graphiques ou textuelles. Tout type d'affichage graphique peut alors être utilisé pour la visualisation des attributs alphanumériques des mesures géographiques. De plus, l'utilisation de la troisième dimension et la possibilité de tourner le cube nous permet de visualiser les relations spatiales entre les mesures et entre les mesures et le membre de la dimension géographique. De plus, l'axe vertical nous permet de visualiser et comprendre les évolutions temporelles du phénomène, le rapport entre les mesures géographiques et un axe de dimension, des clusters de mesures particuliers, etc. En d'autres termes, le cube nous permet de comparer et d'analyser visuellement les mesures géographiques, et la table de pivot de naviguer dans l'hypercube.

### 7.2.2 Un cas d'étude

Pour illustrer GeOlaPivot Table et ses capacités d'analyse, nous décrivons, à travers une maquette d'interface, la navigation dans l'hypercube d'une application multidimensionnelle concernant les maladies infectieuses, comme le SIDA, en Italie. Le schéma de l'application multidimensionnelle est montré en Figure 7.4.

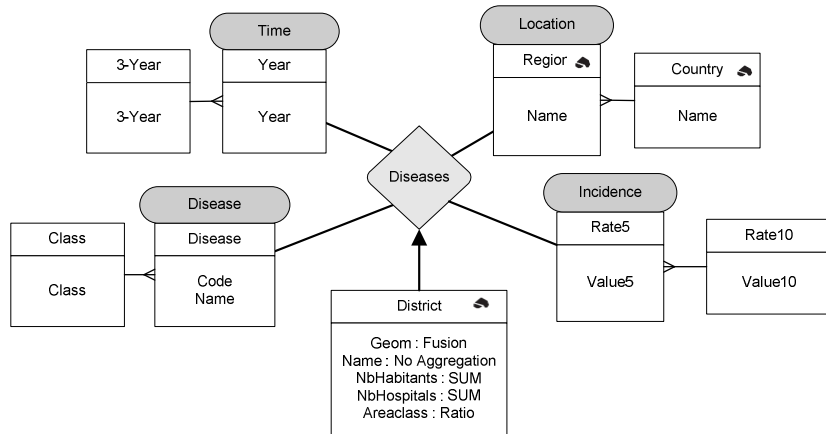


Figure 7.4. Application multidimensionnelle avec une mesure géographique « District » et une dimension géographique « Location ».



Elle présente une dimension géographique « Location » avec une hiérarchie spatiale avec les niveaux « Region » (région) et « Country » (pays) dont les membres sont par exemple « Lombardia », « Italy », etc., une dimension temporelle « Time » avec un niveau reprenant les années (« Year ») et une autre représentant une période de trois années (« 3-Year ») par exemple « 1990-1992 », etc., une dimension représentant les maladies (« Disease ») classée par type de maladie, et enfin une dimension « Rate » qui représente le taux de morts pour 100000 habitants, par exemple « 2,5-3,0 », etc.

Les mesures sont les « Districts » (provinces). Une province est caractérisée par un ensemble d'attributs son nom, la géométrie, le nombre d'hôpitaux et le type (« areaclass »). Les fonctions d'agrégation sont l'union topologique pour la géométrie, la somme du nombre d'habitants et une moyenne pondérée sur la surface pour l'areaclass. Un exemple de données de la table de faits est montré en Table 7.1. Nous avons utilisée un jeu de données extrait du site italien de la santé "Istituto Superiore di Sanità" ([www.iss.it](http://www.iss.it)).

<i>Disease</i>	<i>Time</i>	<i>Incidence</i>	<i>Location</i>	<i>District</i>
AIDS	2001	>6.0	Lombardia	Piacenza
AIDS	2002	>6.0	Lombardia	Brescia, Piacenza
AIDS	2003	>6.0	Lombardia	Lecco, Piacenza
AIDS	2002	>6.0	Emilia- Romagna	Ravenna

Table 7.1. Exemple de la table de faits de l'application multidimensionnelle de Figure 7.4.

La maquette de l'interface que nous avons réalisé modifie et enrichit en fonctionnalités les composants principales de JPivot : le *Cube Navigator* et la table de Pivot.

Le Cube Navigator permet de définir les dimensions et les mesure à utiliser comme lignes ou colonnes de la table de pivot, et de sélectionner des membres pour couper l'hypercube (cf chap 6 § 3.1). En plus de ces fonctionnalités standard, un nouveau champ, appelé Cube Axis a été ajouté. Il permet de sélectionner le niveau représentant l'axe vertical du cube de la GeOlaPivot Table. L'utilisateur peut à tout moment changer une ligne de la table de pivot avec l'axe du cube dans le *Cube Navigator* grâce à l'icône . Enfin, puisque la dimension géographique joue un rôle très important dans la visualisation du cube, elle est associée à l'icône .

La table de pivot a été modifiée afin de présenter dans chaque cellule un cube et huit boutons pour permettre de tourner le cube dans toutes les directions.

Supposons maintenant que l'utilisateur veuille savoir quelles sont les zones d'Italie avec un taux de mortalité supérieure à 6.0 pour le SIDA dans la période 2001-2003. La configuration de l'interface visuelle pour répondre à cette requête est illustrée en Figure 7.5. Nous remarquons qu'en utilisant le champ « Filter » du *Cube Navigator*, l'utilisateur a coupé l'hypercube sur le membre « AIDS » de la dimension « Disease ». De plus, grâce au champ « Cube Axis » du *Cube Navigator*, il affiche le niveau « 3-Year » comme axe vertical du cube. La GeOlaPivot Table contient une seule cellule qui montre un cube avec comme base la représentation cartographique de l'Italie, comme axe vertical le membre « 2001-2003 », et comme point un objet géographique, qui est l'agrégation de plusieurs mesures géographiques détaillées. Cette représentation montre que le SIDA touche les provinces du nord de l'Italie. Qu'est-ce qui caractérise cette zone ? La réponse peut être dans les attributs alphanumériques. Grâce à un simple clic de souris, l'utilisateur visualise alors les attributs pour la mesure agrégée, comme montré en Figure 7.5.



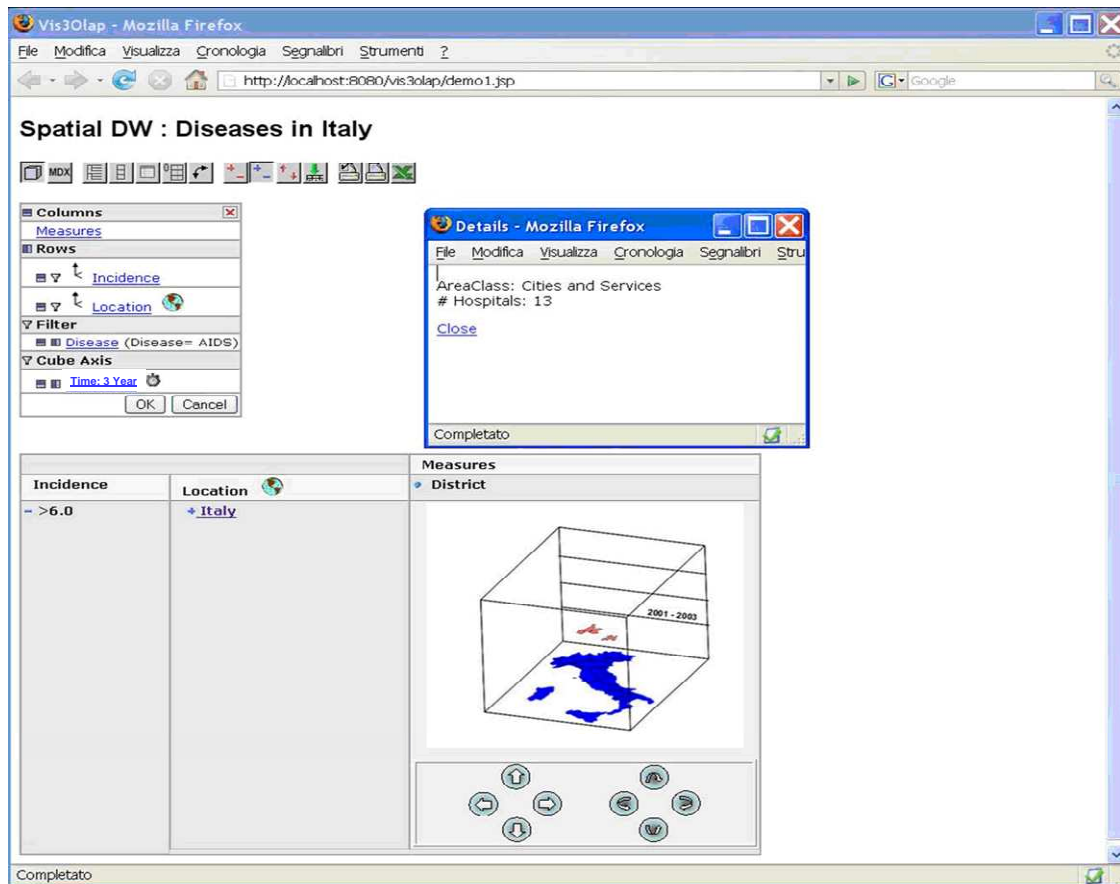


Figure 7.5. Maquette de l'interface qui utilise la GeOlaPivot Table.

Supposons que l'utilisateur veuille avoir un aperçu plus détaillé, et plus particulièrement, qu'il veuille savoir quelles sont les provinces avec un taux de mortalité supérieur à 6.0 touchés par le SIDA pour chaque année. Pour répondre à cette requête il interagit avec la GeOlaPivot Table et le Cube Navigator.

Il applique le drill-down member sur la dimension « Location » en cliquant sur '+' à gauche du membre « Italy » dans la GeOlaPivot Table. En suite, il remplace le niveau « 3-Year » avec « Year » dans le « Cube Axis » du *Cube Navigator*, ce qui correspond à une opération de forage sur la dimension temporelle. Cette action change l'axe du cube dans la GeOlaPivot Table qui montre les provinces pour chaque année. L'interface résultante de ces opérations est montrée en Figure 7.6.

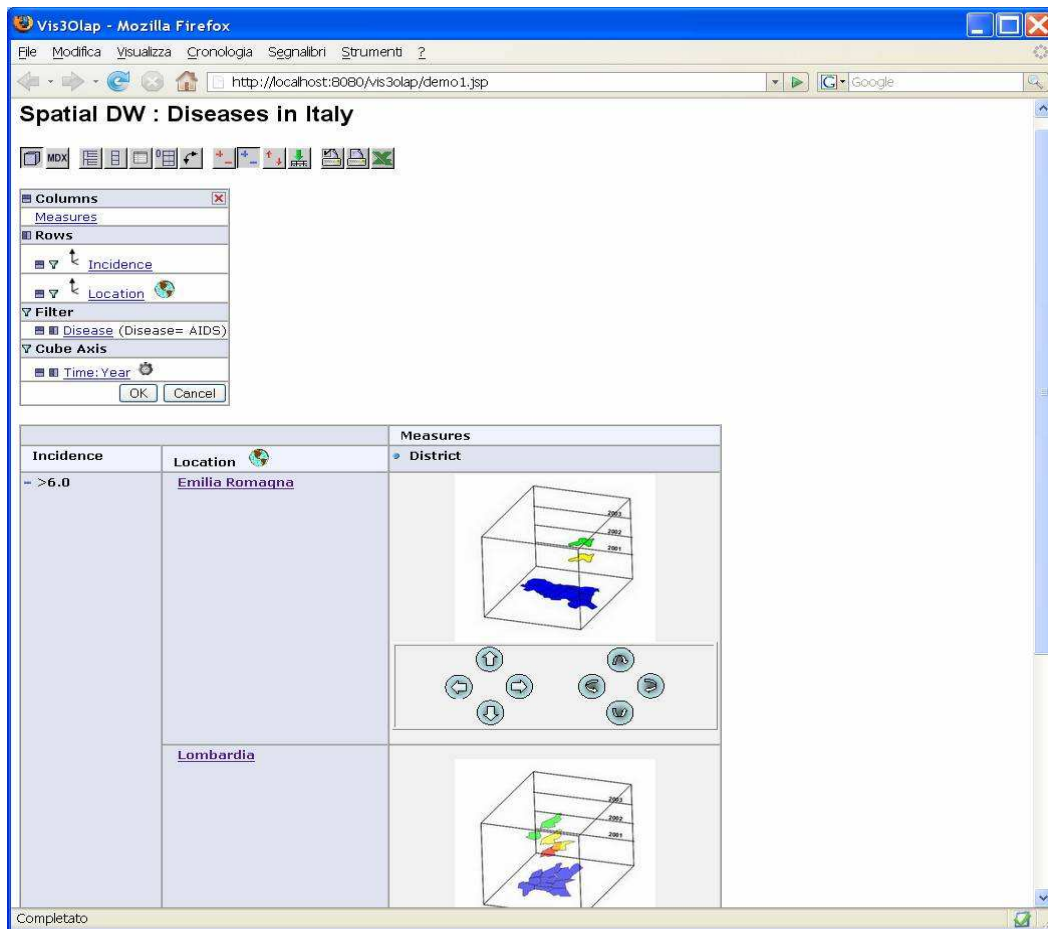


Figure 7.6. Drill down sur la dimension temporelle et la dimension géographique.

Le GeOlaPivot Table permet de « voir » que dans la région Lombardie, trois provinces sont touchées par le SIDA, et en particulier, que la province de Piacenza est toujours affectée par cette maladie de 2001 à 2003. L'information que la même mesure « Piacenza » est associée à différentes combinaisons de membres de dimensions est rendue visible par la troisième dimension du cube et en n'utilisant aucune variable visuelle ou diagramme. De plus, cette configuration de l'interface nous montre que les provinces touchées sont toutes des provinces voisines et que le changement d'année n'implique pas de changement de zone géographique atteinte par la maladie. Le GeOlaPivot Table visualise les relations spatiales entre les mesures et entre les mesures et les membres de dimensions, et les relations non spatiales entre les mesures, tout en gardant une représentation visuelle de la structure de l'application multidimensionnelle et la possibilité de naviguer simplement dans l'hypercube grâce à la seule interaction avec les composantes visuelles de l'interface.

### 7.3 Caractéristiques principales et limitations

La GeOlaPivot Table, en combinant les concepts de table de pivot et de Space-Time Cube, permet de comparer et analyser les mesures géographiques, car elle permet de :

- Visualiser en même temps mesures et dimensions géographiques.
- Visualiser les relations spatiales et aspatiales entre différentes mesures géographiques.
- Visualiser les relations spatiales entre mesures et membres de dimensions.
- Afficher les attributs alphanumériques des objets géographiques.
- Représenter visuellement la structure de l'application multidimensionnelle.
- Rendre accessible les opérateurs OLAP à travers l'interaction avec l'interface visuelle.

Un point critique de la GeOlaPivot Table est lié à la génération de cubes avec les différentes cartes, car les images liées aux différentes mesures géographiques doivent être adaptées à la taille et à la perspective utilisée, en ajoutant un important coût de calcul à celui déjà lourd pour les requêtes multidimensionnelles.

### 7.4 Conclusions

La table de pivot est une des techniques de visualisation et d'interaction la plus adaptée à l'analyse des données multidimensionnelles, car elle permet de comparer les données à différents niveaux de détails, de fournir une visualisation de la structure de l'application multidimensionnelle et de naviguer dans l'hypercube.

Au cas où les mesures sont des données géographiques, la table de pivot n'est pas adaptée à expliciter de façon suffisante les connaissances géospatiales cachées dans l'hypercube. Nous avons introduit un nouveau paradigme de visualisation et d'interaction : la GeOlaPivot Table, qui combine les concepts de table de pivot et Space-Time Cube. La GeOlaPivot Table permet une analyse spatio-multidimensionnelle effective en rendant visible différentes informations comme les relations spatiales et non entre les mesures géographiques, les mesures et les membres, etc. La GeOlaPivot Table représente une technique d'analyse visuelle qui adapte une méthode de géovisualisation à l'analyse des données spatio-multidimensionnelles. Dans ce chapitre, nous avons présenté notre solution, grâce à une maquette d'interface visuelle qui étend celle de JPivot en utilisant un cas d'étude concernant les maladies infectieuses en Italie.



---

## Partie 3 : Conclusions

---



# Chapitre 8 : Bilan et perspectives

## 8.1 Bilan des travaux

Nos travaux concernent l'intégration de l'information géographique dans les entrepôts de données. Nos propositions s'articulent autour de quatre volets :

- un nouveau paradigme pour l'analyse multidimensionnelle des données géographique que nous appelons l'OLAP Géographique,
- GeoCube, un modèle conceptuel et une algèbre pour l'OLAP Géographique,
- un prototype web, fondé sur GeoCube, qui permet l'analyse décisionnelle d'hypercubes géographiques,
- et enfin GeOlaPivot Table, une technique de visualisation et d'interaction pour l'analyse des mesures géographique.

### 8.1.1 Le paradigme OLAP Géographique

Nous introduisons un nouveau paradigme d'analyse multidimensionnelle, l'*OLAP Géographique*, qui prend en compte la composante spatiale et sémantique de l'information géographique et la flexibilité du processus d'analyse spatiale.

Les concepts de mesures et de dimensions spatiales tels que définis dans les modèles SOLAP classiques réduisent généralement l'information géographique à sa composante spatiale. De ce fait, ils sous-exploitent les connaissances encapsulées dans les attributs descriptifs de l'information géographique, de même que les relations spatiales, thématiques et de généralisation cartographique qui pourraient être utiles à l'analyse multidimensionnelle. Nous pensons que l'OLAP Géographique peut être une proposition alternative pour la prise en compte des objets géographiques dans toute leur complexité dans le processus décisionnel. Ainsi, nous définissons une mesure comme un objet géographique appartenant à une hiérarchie. Le sujet d'analyse devient alors un objet géographique caractérisé par plusieurs attributs thématiques et par un attribut spatial. Ces attributs peuvent être utiles au processus décisionnel pour expliquer ou caractériser un ensemble de faits. De plus, un objet géographique peut être en relation avec d'autres objets géographiques et/ou classiques, ce qui permet d'exploiter ces connaissances pour changer la granularité de la mesure. En d'autres termes, cette définition de mesure, contrairement aux modèles classiques, permet la navigation dans une hiérarchie de mesures. Dans les modèles SOLAP, une dimension spatiale est toujours caractérisée par des relations topologiques d'inclusion et/ou d'intersection entre les membres de différents niveaux. De

ce fait, une dimension spatiale ne représente pas explicitement la sémantique des différentes relations hiérarchiques qui peuvent exister entre les objets. L'OLAP Géographique organise les membres des dimensions, représentés par des objets géographiques et/ou classiques, en utilisant trois types différents de hiérarchie : *descriptive*, *spatiale* et *de généralisation*, afin d'exploiter la sémantique des relations hiérarchiques pour la définition des politiques d'agrégation et de navigation. Les hiérarchies descriptives sont des hiérarchies OLAP classiques, construites en utilisant les attributs descriptifs des objets géographiques, et les hiérarchies spatiales sont des hiérarchies SOLAP avec une relation d'inclusion spatiale entre les membres de différents niveaux. Enfin, les hiérarchies de généralisation introduisent la généralisation cartographique dans l'analyse multidimensionnelle. Enfin l'OLAP Géographique redéfinit les opérateurs de navigation multidimensionnelle. Il reformule les opérateurs classiques de forage et de coupe pour prendre en compte les objets géographiques dans le processus d'agrégation et dans la définition des prédicats de coupe. Un opérateur qui permute mesure et dimension, et deux opérateurs qui permettent la navigation dans les hiérarchies des mesures, en utilisant les prédicats SQL ANY et ALL, sont introduits. Afin d'introduire plus de flexibilité dans l'analyse multidimensionnelle, à l'instar de l'analyse spatiale, nous introduisons des opérateurs qui permettent à l'utilisateur de modifier dynamiquement la structure de l'hypercube en créant de nouveaux membres grâce à des opérateurs de transformation de l'analyse spatiale.

### 8.1.2 Le modèle GeoCube

La formalisation des concepts de l'OLAP Géographique soulève un ensemble de problématiques de modélisation multidimensionnelle particulières. Nous présentons ainsi un nouveau modèle multidimensionnel et une algèbre pour l'OLAP Géographique. GeoCube représente les données multidimensionnelles par des objets complexes et/ou géographiques, et ne fait pas de distinction entre dimensions et mesures. Ces objets géographiques sont alors organisés dans des structures hiérarchiques complexes. De cette façon, GeoCube autorise la modélisation des attributs des dimensions, des hiérarchies non-couvrantes et non-strictes, et considère une mesure comme un objet géographique composé par plusieurs attributs simples et dérivés. En utilisant une modélisation symétrique des mesures et des dimensions, GeoCube reformule et étend les opérateurs spatio-multidimensionnels classiques. Les opérateurs de forage et de coupe sont conçus pour prendre en compte les mesures géographiques et pour permettre de couper l'hypercube en utilisant des prédicats qui portent sur les membres des dimensions ou sur les mesures détaillées ou agrégées. L'algèbre de GeoCube introduit l'opérateur Permute qui permet d'intervertir à la volée la mesure et une dimension, et l'opérateur OLAP-Buffer qui modifie dynamiquement la structure de l'hypercube en utilisant l'opérateur d'analyse spatiale de buffer. Si d'un côté GeoCube étend les possibilités d'analyse des modèles SOLAP existants, de l'autre il présente deux limites importantes : il ne prend pas en compte les relations topologiques des hiérarchies spatiales dans le processus d'agrégation, et le type d'attribut (additif, semi-additif et non-additif) dans la définition des fonctions d'agrégation.

### 8.1.3 Le prototype GeWolap

Nous présentons une solution web OLAP-SIG intégrée, *GeWolap*, fondée sur le modèle GeoCube. GeWolap présente toutes les caractéristiques fondamentales d'un



système d'analyse spatio-multidimensionnelle. Il combine le serveur OLAP Mondrian et un client web en s'appuyant sur un SGBD spatial (Oracle). Grâce à Mondrian, GeWolap fournit toutes les fonctionnalités avancées OLAP. Les fonctionnalités de visualisation/interaction cartographique et d'analyse spatiale (i.e. l'ajout de couches, la personnalisation de leurs affichages, etc.) sont implémentées en utilisant les API Java fournies par MapXtreme Java. L'interface visuelle interactive synchronise différentes composantes tabulaire, cartographique et graphique. GeWolap gère les mesures géographiques et complexes, les dimensions géographiques, dans une vision symétrique, et il met en œuvre un ensemble d'opérateurs de l'algèbre de GeoCube. Plus précisément, GeWolap implémente l'opérateur Roll-Up grâce à une extension des mécanismes d'agrégation fournis par Mondrian et à l'utilisation de fonctions définies dans Oracle. L'opérateur Slice sur les membres de dimension, et les opérateurs Buffer et Overlay, qui modifient la structure de l'hypercube, sont aussi fournis par notre prototype. Dans l'état actuel, GeWolap présente des limites du point de vue de la visualisation/interaction et de l'implémentation des opérateurs définis par l'OLAP Géographique. Plus précisément, GeWolap n'adopte pas la technique de visualisation et d'interaction appelée « Multiple Coordinate Windows » qui permettrait la synchronisation de différentes vues des données sur plusieurs fenêtres, et il utilise peu de règles sémiologiques pour la visualisation des mesures numériques. De plus, des techniques particulières pour la pré-agrégation des mesures géographiques doivent être introduites pour améliorer les temps de réponse du système (Stefanovic et al., 2000). Enfin, il nous faudra intégrer des fonctionnalités dans GeWolap telles que l'opérateur Dice, l'extension de l'opérateur Slice aux mesures et la propagation des opérations de coupe à la table des faits, ce qui nous permettra de mettre en œuvre les opérateurs Classify et Specialize comme une navigation entre hypercubes.

#### 8.1.4 GeOlaPivot Table

Le paradigme de visualisation et d'interaction *GeOlaPivot Table* combine les concepts de table de pivot et Space-Time Cube pour l'analyse des mesures géographiques. Les cellules de la table de pivot deviennent des Space-Time Cube dans lesquels l'axe vertical est représenté par une dimension classique, la base du cube est associée à une éventuelle dimension géographique et les valeurs du cube sont les mesures géographiques. Grâce à cette intégration de deux paradigmes d'analyse, GeOlaPivot Table permet de visualiser les relations spatiales et alphanumériques entre les mesures géographiques, et entre les mesures et les membres des dimensions. De plus, GeOlaPivot Table permet à l'utilisateur d'explorer l'hypercube en utilisant le paradigme de visualisation et d'interaction de la table de pivot. GeOlaPivot Table représente alors une technique d'analyse visuelle qui adapte une méthode de géovisualisation à l'analyse des données spatio-multidimensionnelles.

## 8.2 Perspectives

Les perspectives qu'on pourrait poursuivre concernent d'une part GeoCube, et d'autre part la visualisation spatio-multidimensionnelle.

### 8.2.1 Additivité des mesures géographiques

Le problème de l'*additivité des mesures géographiques* reste une problématique ouverte. Comme nous l'avons montré au chapitre 4, dans le cas des mesures géographiques,

le concept d'additivité est fortement influencé par les fonctions d'agrégation spatiales. Les travaux concernant la modélisation des bases de données spatiales (e.g. Voisard et David, 2000), (Rigaux et Sholl, 1998)) et les études portant sur la typologie des attributs des objets géographiques et les opérations de géostatistique (e.g. (Charre et al., 1997)) constituent les bases de nos recherches. Ces travaux formalisent les opérations d'agrégation spatiales et ils fournissent des méthodes pour le choix correct du type d'agrégation à utiliser pour les attributs des objets géographiques. L'idée est alors de combiner ces approches aux solutions OLAP portant sur la sémantique de l'agrégation que nous avons présentées au chapitre 1. Il est important de souligner que pour une agrégation correcte des mesures géographiques, de nombreuses problématiques liées à la topologie des mesures (Pedersen et Tryfona, 2001) doivent aussi être résolues.

### 8.2.2 Hiérarchies de généralisation

La prise en compte de la complexité des *hiérarchies de généralisation* et la définition de politiques d'agrégation ad hoc au sein d'un modèle et d'un système OLAP sont deux problématiques ouvertes. La modélisation des relations multi-association et leur influence dans le processus d'agrégation constituent un défi important. Comment peut-on introduire dans un modèle multidimensionnel les relations hiérarchiques dérivant de la généralisation cartographique ? Comment peut-on lier l'information associée aux mesures dans les différents niveaux des hiérarchies de généralisation à travers des fonctions d'agrégation ? Comment prendre en compte le degré d'imprécision qui est implicitement associé à ces hiérarchies ? Une piste intéressante est fournie par les modèles de données spatiales multi-granulaires, car ils définissent des fonctions de conversion pour représenter la même information à différentes granularités (Camossi et al., 2006). Les modèles multidimensionnels qui utilisent la logique floue pour la représentation des mesures et des hiérarchies (Delgado, et al., 2004) nous semblent également être des solutions exploitables. Dans ce travail, nous nous sommes toujours intéressé au changement d'échelle dans la définition des hiérarchies de généralisation. Dans un contexte plus large, les opérateurs de généralisation peuvent être utilisés pour obtenir des cartes secondaires sans changer forcément l'échelle topographique (Model-Oriented Generalization) (Weibel et Dutton, 2001). Une analyse des problématiques pour ce type de généralisation est nécessaire pour un modèle multidimensionnel qui inclut la généralisation des cartes dans tous ces aspects.

### 8.2.3 Composante temporelle de l'information géographique

L'introduction des aspects temporels dans l'OLAP est une problématique importante (Body et al, 2003). L'information géographique présente souvent une composante temporelle. (Miquel et al., 2002) illustrent ce problème en utilisant un cas d'étude dans le domaine de la foresterie. Un inventaire forestier consiste à partitionner la surface de la forêt étudiée en zones présentant des caractéristiques forestières homogènes (essence, âge, densité, hauteur, etc.). Une application spatio-multidimensionnelle pour l'analyse des zones de la forêt présente comme mesures la surface de la zone ainsi que 5 dimensions : Essence, Age, Densité, Temps et Découpage, cette dernière étant la dimension spatiale qui décrit une zone. L'hétérogénéité des données géométriques lorsque le découpage de l'espace évolue au cours du temps, et l'évolution des données descriptives en fonction de la législation, du mode d'acquisition, de la variation des domaines de valeurs des attributs et de leur codage, caractérisent les membres de la dimension spatiale. La gestion de l'évolution temporelle

des structures et des instances des dimensions spatiales devient alors une caractéristique fondamentale des applications spatio-multidimensionnelles. L'intégration de la composante temporelle dans le SOLAP a été l'objet de différents travaux, qui, pour la majorité, se concentrent sur les modèles logiques (Miquel et al., 2002), (Pestana et al., 2005). En s'appuyant sur ces travaux, il serait souhaitable d'intégrer la composante temporelle dans GeoCube et dans notre prototype.

#### 8.2.4 Mesures continues

L'introduction du modèle matriciel dans un modèle OLAP, comme montré par (Ahmed et Miquel, 2005), est une voie prometteuse. Pour cela, il est possible d'étendre notre approche (formelle et pratique) pour permettre une analyse multidimensionnelle des données géographiques continues (mesures continues), et améliorer les capacités de manipulation de rasters des SIG classiques (Gutierrez et Baumman, 2007). L'adaptation et l'intégration des méthodes de la Map Algebra, en tant que fonctions d'agrégation dans le modèle formel et dans GeWolap représentent des défis majeurs. Comment peut-on modéliser la continuité spatiale au sein de GeoCube ? Comment définir l'agrégation des objets continus dans le serveur OLAP ? Comment visualiser et surtout analyser les mesures continues ? L'adoption d'un SGDB, qui permet de gérer les rasters, comme Oracle qui intègre parmi les types de données classiques le type GEORASTER, nous semble une solution possible pour étendre les mécanismes d'agrégation de GeWolap pour la prise en compte des données continues, et ainsi répondre aux questions posées ci-dessus. En ce qui concerne la visualisation, l'intégration des cartes continues dans les cellules de la table de pivot est selon nous une approche satisfaisante. Elle serait effectivement une extension et une adaptation de la technique de géovisualisation MultiMaps qui permet de comparer visuellement les différentes données thématiques des données géographiques.

#### 8.2.5 Géovisualisation et OLAP Géographique

Une étude globale qui vise à définir les *techniques de géovisualisation* les plus adéquates à l'analyse OLAP Géographique, afin de les intégrer dans un seul système d'aide à la décision spatiale, est souhaitable. Cette problématique soulève de nombreuses questions car les modèles de données et d'interaction utilisés dans les systèmes des Exploratory Spatial Data Analysis (ESDA) et du SOLAP sont très différents. Les systèmes ESDA n'utilisent pas de modèle multidimensionnel, mais reposent sur une organisation en couches. De plus, dans les systèmes ESDA, les chemins de l'analyse sont flexibles, en d'autres termes, ils permettent à l'utilisateur de définir ses propres axes d'analyse à travers l'intégration de différentes méthodes d'analyse spatiale. Au contraire, dans les solutions SOLAP, le modèle de données spatio-multidimensionnelles définit les analyses spatiales possibles (requêtes spatio-multidimensionnelles) et les techniques de visualisation (cartes interactives), en enfermant le processus décisionnel dans un schéma d'analyse rigide. Ainsi, une adaptation des méthodes de géovisualisation, comme nous l'avons proposée avec GeOlaPivot Table, et/ou la définition de structures de données particulières dédiées à l'intégration et à la collaboration entre les deux systèmes (Voss et al., 2004), (Han et al., 1997) semblent des pistes envisageables.

Nous pensons que la définition de règles sémiologiques pour la visualisation des données spatio-multidimensionnelles en fonction du type de dimension, de mesure, etc. est

une étape fondamentale vers la conception d'un système pour l'analyse des données spatio-multidimensionnelles (Soumahoro, 2005). Le système devrait pouvoir sélectionner automatiquement les variables visuelles (couleurs, taille, etc.) et les représentations graphiques les plus adaptées pour la visualisation des mesures en fonctions du type (qualitative, quantitative) et du nombre de mesures, du type d'implantation spatiale (point, ligne, polygone) et du type d'agrégation utilisée (somme, moyenne, count, etc.). Une visualisation « simplifiée » des membres spatiaux et des mesures (spatiales et/ou numériques) à travers l'utilisation des chorèmes, peut selon nous améliorer le pouvoir expressif des cartes interactives des solutions SOLAP. Les chorèmes fournissent un langage visuel pour la description des phénomènes spatiaux d'aide au processus cognitif. Ils synthétisent les aspects spatiaux et thématiques des objets et des phénomènes géographiques à travers des représentations graphiques particulières.

### **8.2.6 L'OLAP Géographique Mobile**

Nous pensons enfin que la définition d'un système SOLAP pour dispositifs mobiles est une piste de recherche intéressante et fascinante. (Bédard et al., 2006) montrent que l'utilisation d'éléments multimédia comme les photos etc. aident l'utilisateur dans son analyse multidimensionnelle. A travers un système SOLAP mobile, l'utilisateur ne doit plus utiliser de photos pour positionner son analyse dans le vrai contexte environnemental, mais peut simplement se promener sur le terrain, prendre les mesures, les comparer à la volée aux données historiques des entrepôts. Nous pensons qu'un enjeu majeur pour l'OLAP Géographique mobile reste l'adaptation du paradigme de visualisation et d'interaction typique du SOLAP (cartes interactives, tables de pivot et graphiques) à la taille réduite des écrans des dispositifs mobiles. Quel type de cartes utiliser, comment adapter la table de pivot ? Comment implémenter le « Multiple Coordinate Windows » ? Comment intégrer la position de l'utilisateur dans les requêtes (formulation et visualisation) spatio-multidimensionnelles ? etc. Ces interrogations constituent des pistes de recherche très prometteuses.

L'introduction des données géographiques dans les solutions informatiques d'aide à la décision est une thématique de grande actualité. Le "géo-décisionnel" offre des solutions nouvelles pour des applications, comme le géomarketing, l'analyse environnementale, etc. L'intégration des données géospatiales dans les entrepôts de données et l'analyse en ligne suscite un fort intérêt de la part du monde académique et industriel. Notre travail de thèse s'inscrit alors dans ce contexte fascinant de recherche pour lequel de nombreux et importants défis restent encore à résoudre.

# Références

(Abelló, et al., 2006) ABELLÓ Alberto, SAMOS José et SALTOR Félix. YAM<sup>2</sup>: A Multidimensional Conceptual Model Extending UML. *Information Systems*, 2006, Vol. 3, n° 6, 541-567 p.

(Agrawal et al., 1997) AGRAWAL Rakesh, GUPTA Ashish et SARAWAGI Sunita. Modeling Multidimensional Databases. In : GRAY W. A. et LARSON Per-Åke. 13th International Conference on Data Engineering, 7-11 Avril 1997, Birmingham, U.K. IEEE Computer Society, 1997, 232-243 p.

(Ahmed et Miquel, 2005) AHMED Taher et MIQUEL Maryvonne. Multidimensional Structures Dedicated to Continuous Spatiotemporal Phenomena. In : JACKSON Mike, NELSON David et STIRK Sue. 22th British National Conference on Databases, 5-7 Juillet 2005, Sunderland, UK. Berlin Heidelberg : Springer, 2005, 29-40 p. (Lecture Notes in Computer Science 3567)

(Alter, 1980) ALTER Steven. *Decision Support System: Current Practices and Continuing Challenges*. Massachusetts: Addison-Wesley Publishing Co., 1980, 316 p.

(Andrienko et al., 2003) ANDRIENKO Natalia V., ANDRIENKO Gennady L. et GATALSKY Peter. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages and Computing*, 2003, Vol. 14, n° 6, 503-541 p.

(Andrienko et Andrienko, 2003) ANDRIENKO Natalie et ANDRIENKO Gennady. Informed spatial decisions through coordinated views. *Information Visualization*, 2003, Vol. 2, n° 4, 270-285 p.

(Armstrong et Densham, 1990) ARMSTRONG Marc P. et Densham Paul J. Database organization strategies for spatial decision support systems. *International Journal of Geographical Information Systems*, 1990, Vol. 4, n° 1, 3-20 p.

(Bauer, et al., 2000) BAUER Andreas, HÜMMER Wolfgang et LEHNER Wolfgang. An Alternative Relational OLAP Modeling Approach. In : KAMBAYASHI Yahiko, MOHANIA Mukesh K. et TJOA A. Min. *Data Warehousing and Knowledge Discovery, Second International Conference*, 4-6 Septembre, 2000, Londres, UK. Berlin Heidelberg : Springer, 2000, 189-198 p. (Lecture Notes in Computer Science 1874)

(Bédard, 1997) BÉDARD Yvan. Spatial OLAP. 2ème Forum annuel sur la R-D, Géomatique VI: Un monde accessible, 13-14 Novembre ,1997, Montréal

(Bédard et al, 2001) BÉDARD Yvan, MERRETT T. H. et HAN Jiawei. Fundaments of Spatial Data Warehousing for Geographic Knowledge Discovery. In : MILLER Harvey J. et HAN Jiawei. Geographic Data Mining and Knowledge Discovery. Londres : Taylor & Francis, 2001, 53-73, p.

(Bédard et al 2002) BÉDARD Yvan, GOSSELIN Pierre, RIVEST Sonia, PROULX Marie-Josée, NADEAU Martin, LEBEL Germain et GAGNON Marie-France. Integrating GIS Components with Knowledge Discovery Technology for Environmental Health Decision Support. International Journal of Medical Informatics, 2002, Vol. 70, n° 1, 79-94 p.

(Bédard, et al., 2005) BÉDARD Yvan, PROULX Marie-Josée et RIVEST Sonia. 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, 2005, 1-20 p.

(Bédard, et al. 2006) BÉDARD Yvan, PROULX Marie-Josée, RIVEST Sonia et BADARD Thierry. Merging Hypermedia GIS with Spatial On-Line Analytical Processing: Towards Hypermedia SOLAP. In: STEFANAKIS Emmanuel, PETERSON Michael P. et DELIS Vasilis. Geographic Hypermedia: Concepts and Systems. Berlin : Springer, 2006, 167-185 p. (Lecture Notes in Geoinformation and Cartography)

(Bédard et al., 2007) BÉDARD Yvan, RIVEST Sonia et PROULX Marie-Josée. Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective. In : Wrembel Robert et Koncilia Christian. Data Warehouses and OLAP : Concepts, Architectures and Solutions. Londres, UK : IRM Press (Idea Group), 2007, 298-319 p.

(Bertin et Bonin, 1992) BERTIN Jacques et BONIN Serge. La graphique et le traitement graphique de l'information. Paris : Flammarion, 1992, 277 p. (Nouvelle bibliothèque scientifique)

(Bimonte et al., 2005) BIMONTE Sandro, TCHOUNIKINE Anne et MIQUEL Maryvonne. Towards a spatial multidimensional model. In : SONG Il-Yeol et TRUJILLO Juan. ACM 8th International Workshop on Data Warehousing and OLAP, 4-5 Novembre, 2005Bremen, Allemagne. New York, NY, USA : ACM press, 2005, 39-46 p.

(Bimonte et al., 2006) BIMONTE Sandro, TCHOUNIKINE Anne et MIQUEL Maryvonne. GeoCube, a Multidimensional Model and Navigation Operators Handling Complex Measures: Application in Spatial OLAP. In : YAKNHO Tatyana M. et NEUHOLD Erich J. Advances in Information Systems, 4th International Conference, 18-20 Octobre, 2006, Izmir, Turquie. Berlin-Heidelberg : Springer, 2006, 100-109 p. (Lecture Notes in Computer Science 4243)

(Bimonte et al., 2006b) BIMONTE Sandro, WEHRLE Pascal, TCHOUNIKINE Anne et MIQUEL Maryvonne. GeWolap: A Web Based Spatial OLAP Proposal. In : MEERSMAN Robert, TARI Zahir et HERRERO Pilar. Workshop on Semantic-Based

Geographical Information Systems, 29-30 Octobre, 2006 Montpellier, France. Berlin-Heidelberg : Springer, 2006, 1596-1605 p. (Lecture Notes in Computer Science 4278)

(Bimonte et al., 2006c) BIMONTE Sandro, DI MARTINO Sergio, FERRUCCI Filomena et TCHOUNIKINE Anne. GeOlaPivot Table: a Visualization Paradigm for SOLAP Solutions. In Visual Languages and Computing Workshop, 30 Août – 1 Septembre, 2006, Grand Canyon, USA. K. S. Institute, 2006, 181-186 p.

(Bimonte et al., 2006d) BIMONTE Sandro, TCHOUNIKINE Anne, MIQUEL Maryvonne, LAURINI Robert et AHMED Taher. Spatial Online Analytical Processing for Environmental Data. In: CORILA. Scientific Research and Safeguarding of Venice. Research Programme 2004-2006, Vol. 4, 2005 results. Venise, Italie: CORILA, 2006, 393-400 p.

(Bimonte et al., 2007) BIMONTE Sandro, TCHOUNIKINE Anne et MIQUEL Maryvonne. Spatial OLAP: Open Issues and a Web Based Prototype. In : WACHOWICZ Monica et BODUM Lars. 10th AGILE International Conference on Geographic Information Science, 8-11 Mai 2007, Aalborg, Danemark

(Bimonte et al., 2007a) BIMONTE Sandro, DI MARTINO Sergio, FERRUCCI Filomena et TCHOUNIKINE Anne. Supporting Geographical Measures Through A New Visualization Metaphor In Spatial OLAP. In : CARDOSO Jorge, CORDEIRO Jose et FILIPE Joaquim. 9th International Conference on Enterprise Information Systems, 12-16 Juin, 2007, Funchal, Madeira, Portugal. INSTICC, 2007, 19-26 p.

(Bimonte et al., 2007b) BIMONTE Sandro, TCHOUNIKINE Anne et MIQUEL Maryvonne et LAURINI Robert. Vers l'intégration de l'analyse spatiale et multidimensionnelle. In Colloque International de GÉomatique et d'Analyse Spatiale, 18 - 20 juin 2007, Clermont-Ferrand, France

(Blaschka et al., 1998) BLASCHKA Markus, SAPIA Carsten, HÖFLING Gabriele et DINTER Barbara. Finding your way through multidimensional data models. In : WAGNER Roland. 9nd International Workshop on Database and Expert Systems Applications, 24-28 Aout, 1998, Vienne, Autriche. Washington, DC, USA : IEEE Computer Society, 1998, 198-203 p.

(Bobzien et al., 2006) BOBZIEN, M., BURGHARDT, D., PETZOLD, I., NEUN, M et WEIBEL Robert. Multi-Representation Databases With Explicitly Modelled Intra-Resolution, Inter-Resolution and Update Relations. In AutoCarto American Congress on Surveying and Mapping, 26-28 Juin, 2006, Vancouver, USA.

(Body et al., 2003) BODY Mathurin, MIQUEL Maryvonne, BÉDARD Yvan, TCHOUNIKINE Anne. Handling Evolutions in Multidimensional Structures. In : DAYAL Umeshwar, RAMAMRITHAM Krithi, VIJAYARAMAN T. M.. 19th International Conference on Data Engineering, 5-8 Mars, 2003, Bangalore, Inde. IEEE Computer Society, 2003, 581- p.

(Burdick et al., 2005) BURDICK Douglas, DESHPANDE Prasad, JAYRAM T. S., RAMAKRISHNAN Raghu et VAITHYANATHAN Shivakumar. OLAP Over Uncertain and Imprecise Data. *International Journal on Very Large Data Bases*, 2005, Vol. 16, n° 1, 123-144 p.

(Cabibbo et Torlone, 1998) CABIBBO Luca et TORLONE Riccardo. A logical approach to multidimensional databases. In : SCHEK Hans-Jörg, SALTOR Fèlix, RAMOS Isidro et ALONSO Gustavo. 6th International Conference on Extending Database Technology, 23-27 Mars 1998, Valencia, Espagne. Berlin Heidelberg: Springer-Verlag, 1998, 183–197 p. (Lecture Notes in Computer Science 1377)

(Charre et al., 1997) CHARRE Joelle, DUMOLARD Pierre et LE BERRE I. *Initiation aux pratiques statistiques en géographie*. Paris : Masson, 1997, 185 p.

(Camossi et al., 2006) CAMOSSO Elena, BERTOLOTTA Michela et BERTINO Elisa. A multigranular object-oriented framework supporting spatio-temporal granularity conversions. *International Journal of Geographical Information Science*, 2006 Vol. 20, n° 5, 511-534 p.

(Caron, 1998) CARON Pierre-Yves, *Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle*. Rapport de DEA. DEA Informatique, Laval : Université Laval, Canada, 1998, 129 p.

(Colonese et al., 2005) COLONESE Giovanni, MANHÃES Rodrigo S., MONTENEGRO Sahudy G., CARVALHO Rogério et TANAKA Asterio K. PostGeoOlap: an Open-Source Tool for Decision Support. In: 2nd Simpósio Brasileiro de Sistemas de Informação, 26-30 Septembre, 2005, Florianópolis, Brésil

(Compieta et Al, 2007) COMPIETA Paul, DI MARTINO Sergio, BERTOLOTTA Michela, FERRUCCI Filomena et KECHADI Tahar. Exploratory spatio-temporal data mining and visualization. Elsevier : *Journal of Visual Languages and Computing*, 2007, Vol. 18, n° 3, à paraître

(Couclelis, 1992). COUCLELIS Helen. People manipulate objects (but cultivate fields) : beyond the raster-vector debate in GIS. In : FRANK Andrew, CAMPARI Irene et FORMENTINI Ubaldo. *Theories and methods of spatio-temporal reasoning in geographic space : International Conference Gis-From Space to Territory : Theories and Methods*, 1992, Berlin: Springer Verlag, 65-77 p.

(Damiani et Spaccapietra, 2006) DAMIANI Maria Luisa et SPACCAPIETRA Stefano. Spatial Data Warehouse Modeling. In : DARMONT Jérôme et BOUSSAID Omar. *Processing and Managing Complex Data for Decision Support*. Hershey, PA, USA : IDEA Group Publishing, 2006, 1-27 p.

(Degrene et Salgé, 1997) DENEGRE Jean et SALGE François. *Les systèmes d'information géographique*. 2nd Ed. Paris : Presses Universitaires de France, 2004, 128 p. (Que sais-je?)



(Delgado, et al., 2004) DELGADO Miguel, MOLINA Carlos, SANCHEZ Daniel, VILA Amparo et RODRIGUEZ-ARIZA Lazaro. A fuzzy multidimensional model for supporting imprecision in OLAP. In : IEEE International Conference on Fuzzy Systems, 25 Juillet, Budapest, Hongrie, 1331-1336 p.

(Dyreson, 1996) DYRESON Curtis E.. Information Retrieval from an Incomplete Data Cube. In : VIJAYARAMAN M., BUCHMANN Alejandro P., MOHAN C. et SARDA Nandlal L., 22th International Conference on Very Large Data Bases, 3-6 Septembre, 1996, Mumbai (Bombay), Inde. San Francisco : Morgan Kaufmann, 1996, 532-543 p.

(Dykes et al., 2005) DYKES Jason, MACEACHREN Alan M., et KRAAK Menno-Jan. Exploring Geovisualization. Amsterdam, Hollande: Elsevier, 2005, 710 p.

(Dykes et Mountain, 2003) DYKES Jason et MOUNTAIN David. Seeking structure in records of spatio-temporal behaviour: visualization issues, efforts and applications. Computational Statistics and Data Analysis, 2003, Vol. 43, n° 4, 581-603 p.

(Eder et Koncilia, 2001) EDER Johann et KONCILIA Christian. Changes of Dimension Data in Temporal Data Warehouses. In : KAMBAYASHI Yahiko, WINIWARTER Werner et ARIKAWA Masatoshi. 3th International Conference Data Warehousing and Knowledge Discovery, 5-7 Septembre, 2001, Munich, Allemagne. Berlin Heidelberg : Springer, 2001, 284-293 p. (Lecture Notes in Computer Science 2114)

(Eick, 2000) EICK Stephen G. Visualizing multi-dimensional data. SIGGRAPH Computer Graphics, 2000, Vol. 34, n° 4, 61-67 p.

(Elmasri et Navathe, 2003) ELMASRI Ramez et NAVATHE Shamkant B.. Fundamentals of Database Systems. 4th Ed. Boston : Addison Wesley, 2003, 1009 p.

(ESRI, 2007) ESRI. Olap for ArcGIS. [en ligne]. Disponible sur : <<http://www.esri.com/software/arcgis/extensions/olap/index.html>> (consulté le 1.09.2007).

(Ferreira et al., 2001) FERREIRA Ana C., CAMPOS Maria L. et TANAKA Astério. An Architecture for Spatial and Dimensional Analysis Integration. In : Callaos Nagib, Silva Ivan, Molero Jorge. World Multiconference on Systemics, Cybernetics and Informatics, 22-25 Juillet, 2001, Orlando, Florida, USA. IIIS, ISAS-SCI, 2001 392-395 p.

(Franklin, 1992) FRANKLIN Carl. An Introduction to Geographic Information Systems: Linking Maps to databases. Database, 1992, Vol. 15, n° 2, 13-21 p.

(Fidalgo et al. 2004) FIDALGO Robson N., TIMES Valeria C., SILVA Joel et SOUZA Fernando F. GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas. In : KAMBAYASHI Yahiko, MOHANIA Mukesh K. et WÖß Wolfram. 6th International Conference on Data Warehousing and Knowledge Discovery, 1-3 Septembre, 2004, Saragosse, Espagne. Berlin Heidelberg : Springer, 2004, 26-37 p. (Lecture Notes in Computer Science 3181)

(Galigeo, 2007) GALIGEO SOLUTIONS. Webigeo. [en ligne]. Disponible sur : <<http://www.galigeo.com/?/Produits/WebiGeo/Description-de-WEBIGEO>> (consulté le 1.09.2007).

(Gatalsky et AL., 2004) GATALSKY Peter, ANDRIENKO Nataly et ANDRIENKO Gennady. Interactive analysis of event data using space-time cube. In : BANISSI Eric. 8th International Conference on Information Visualization, Juillet 2004, Londres. Los Alamitos : IEEE Computer Society, 2004, 145-152 p.

(Goodchild et Densham, 1990) GOODCHILD Michael Frank et DENSHAM Paul J. Spatial Decision Support Systems: Scientific Report for the Specialist Meeting, TR-90-5. Santa Barbara, California : National Center for Geographic Information and Analysis, 1990, 124 p.

(Gutierrez et Baumman, 2007) GUTIERREZ Angelica Garcia et BAUMANN Peter. Modeling Fundamental Geo-Raster: Operations with Array Algebra. TR n° 7. Jacobsen University, Breme, Allemagne, 2007, 41 p.

(Hägerstrand, 1070) HÄGERSTRAND Torsten. What About People in Regional Science? Papers of the Regional Science Association, 1970, Vol. 24, 7-21 p.

(Haining, 2003), HAINING Robert. Spatial Data Analysis: Theory and Practice. New York : Cambridge University Press, 2003, 452 p.

(Han et al., 1997) HAN Jiawei, KOPERSKI Krzysztof et STEFANOVIC Nebojsa. GeoMiner: A System Prototype for Spatial Data Mining. In : PECKHAM Joan. ACM SIGMOD International Conference on Management of Data, 13-15 Mai, 1997, Tucson, Arizona, USA. New York : ACM Press, 1997, 553-556 p.

(Harinarayan et al., 1996) HARINARAYAN Venky, RAJARAMAN Anand et ULLMAN Jeffrey D. Implementing Data Cubes Efficiently. In : JAGADISH H. V. et MUMICK Inderpal Singh. ACM SIGMOD International Conference on Management of Data, 4-6 Juin, 1996, Montreal, Quebec, Canada. New York : ACM Press, 1996, 205-216 p.

(Hernández et al., 2005) HERNANDEZ Vera, VOSS Angi, GÖHRING Wolf et HOPMANN Cornelio. Sustainable decision support by the use of multi-level and multi-criteria spatial analysis on the Nicaragua Development Gateway. In : From pharaohs to geoinformatics Proceedings of FIG Working Week 2005 and 8th International Conference on the Global Spatial Data Infrastructure, 16-21 Avril 2005, Le Caire, Egypte.

(Hernández, 2006) HERNÁNDEZ Ernst Vera. Affordable web-based spatio-temporal applications for Ad-Hoc decisions. In : MEERSMAN Robert, TARI Zahir et HERRERO Pilar. On the move to meaningful internet systems 2006. Vol.2 : OTM 2006 workshops, SeBGIS 2006, 29 Octobre – 3 Novembre, 2006, Montpellier, France. Berlin: Springer, 2006, 1606-1615 p. (Lecture Notes in Computer Science 4278)

(Horner et al., 2004) HORNER John, SONG Il-Yeol et CHEN Peter P. An Analysis of Additivity in OLAP Systems. In : SONG Il-Yeol et DAVIS Karen C. 7th ACM

international workshop on Data warehousing and OLAP, 12-13 Novembre, 2004, Washington, DC, USA. New York, NY, USA : ACM Press, 2004, 83-91 p.

(Hurtado et al., 1999) HURTADO Carlos A., MENDELZON Alberto O. et VAISMAN Alejandro A. Maintaining Data Cubes under Dimension Updates. In : 15th International Conference on Data Engineering, 23-26 Mars 1999, Sydney, Australie. Los Alamitos, California : IEEE Computer Society Press, 1999, 346-355 p.

(Hüsemann et al., 2000) HÜSEMANN Bodo, LECHTENBÖRGER Jens et VOSSSEN Gottfried. Conceptual data warehouse modeling. In : JEUSFELD Manfred A., SHU Hua, STAUDT Martin et VOSSSEN Gottfried. econd Intl. Workshop on Design and Management of Data Warehouses, June 5-6, 2000, Stockholm, Suède. CEUR-WS.org, 2000, Vol. 28, 6.

(ICA, 1973) INTERNATIONAL CARTOGRAPHIC ASSOCIATION. Multilingual dictionary of technical temrs in cartography. Wiesbaden : Franz Steiner, 1973, 573 p.

(Inmon, 1996) INMON W.H. Building the Data Warehouse. 2nd Ed. New York : Wiley, 1996, 401 p.

(ISO, 2002) INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. SQL multimedia and application packages—part 3: Spatial. FCD 13249-3I, 2002

(Jensen et al. 2004) JENSEN Christian S., KLIGYS Augustas, PEDERSEN Torben Bach et TIMKO Igor. Multidimensional data modeling for location-based services. International Journal on Very Large Data Bases, 2004, Vol.13, n° 1, 1-21 p.

(JMap, 2007) KHEOPS. [en ligne]. Disponible sur : <<http://www.kheops-tech.com/fr/jmap/solap.jsp>> (consulté le 1.09.2007).

(JPivot, 2007) JPIVOT. [en ligne]. Disponible sur : <<http://jpivot.sourceforge.net/>> (consulté le 1.09.2007).

(Keenan, 1996) KEENAN Peter. Using a GIS as a DSS Generator. In : DARZENTAS John, DARZENTAS Jenny et SPYROU Thomas. Perspectives on Decision Support System. Grèce : University of the Aegean, 1996, 33-40 p.

(Kelly, 1997) KELLY Sean. Data warehousing in action. Chichester, West Sussex, England : John Wiley & Sons, 1997, 334 p.

(Kimball, 1996) KIMBALL Ralph. The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses. New York : John Wiley & Sons, 1996, 374 p.

(Köbben, 2001) KÖBBEN Barend. Publishing maps on the Web. In : KRAAK Jan-Menno et BROWN Allan. Web Cartography. Londres, UK : CRC, 2001, 213 p.

(Kouba, et al., 2000) KOUBA Zdenek, MATOUSEK Kamil et MIKSOVSKÝ Petr. On Data Warehouse and GIS Integration. In : IBRAHIM Mohamed T., KÜNG Josef et REVELL Norman. 11th International Conference on Database and Expert Systems

Applications, 04-08 Septembre, 2000, Londres, UK. Londres, UK : Springer, 2000 (Lecture Notes in Computer Science 1873) 604-613 p.

(Kraak, 2003) KRAAK Meno-Jan. The space-time cube revisited from a geovisualization perspective. In : 21th International Cartographic Conference : Cartographic Renaissance, 10-16 Aout 2003, Durban, Sud Afrique, 1988-1996 p.

(Kraak et Koussoulakou, 2004) KRAAK Menno-Jan et KOUSSOULAKOU Alexandra. A Visualization Environment for the Space-Time-Cube In : FISCHER Peter F. Developments in Spatial Data Handling. Berlin Heidelberg : Springer, 2004, 189-200 p.

(Laurini et Milleret-Raffort, 1993) LAURINI Robert et MILLERET-RAFFORT Françoise. Les Bases de Données en Géomatique. Paris : HERMES, 1993, 340 p (Collection Géomatique)

(Lehner, 1998) Wolfgang Lehner. Modelling Large Scale OLAP Scenarios. In : SCHEK Hans-Jörg, SALTOR Fèlix, RAMOS Isidro et ALONSO Gustavo. 6th International Conference on Extending Database Technology, 23-27 Mars 1998, Valencia, Espagne. Berlin Heidelberg: Springer-Verlag, 1998, 153-167 p. (Lecture Notes in Computer Science 1377)

(Lenz et Shoshani, 1997) LENZ Hans-Joachim et SHOSHANI Arie. Summarizability in OLAP and Statistical Databases. In: IOANNIDIS Yannis E., HANSEN David M.. 9th International Conference on Scientific and Statistical Databases Management, 11-13 Aout, 1997, Olympia, Washington, USA. Los Angeles, CA : IEEE Computer Society 1997, 132-143 p.

(Li et al., 2003) LI Ying, CHEN Ying et RAO Fangyan. The Approach for Data Warehouse to Answering Spatial OLAP Queries. In : LIU Jiming, CHEUNG Yiu-ming et YIN Hujun. Intelligent Data Engineering and Automated Learning, 4th International Conference, 21-23 Mars, 2003, Hong Kong, Chine. Berlin Heidelberg : Springer, 2003, 270-277 p. (Lecture Notes in Computer Science 2690)

(Li et Wang, 1996) LI Chang et WANG Xiaoyang Sean. A Data Model for Supporting On-Line Analytical Processing. In : 5th International Conference on Information and Knowledge Management, 12-16 Novembre, 1996, Rockville, Maryland, USA. New York : ACM 1996, 81-88 p.

(Longley, et al., 2001) LONGLEY Paul, GOODCHILD Michael, MAGUIRE David et RHIND David. Geographic Information Systems and Science. New York : John Wiley & Sons, 2001, 517 p.

(Lujian-Mora et al., 2002) LUJAN-MORA Sergio, TRUJILLO Juan et SONG II-Yeol. Multidimensional modeling with UML package diagrams. In : SPACCAPIETRA Stefano, MARCH Salvatore T. et KAMBAYASHI Yahiko. 21th International Conference on Conceptual Modeling, 7-11 Octobre, 2002, Tampere, Finlande. Berlin : Springer, 2002, 199-213 p. (Lecture Notes in Computer Science 2503)

(MacEachren, 2001) MACEACHREN Alan M. An evolving cognitive-semiotic approach to geographic visualization and knowledge construction. *Information Design Journal*, 2001, Vol. 10, n° 1, 26-36 p.

(MacEachren, et al., 2004) MACEACHREN Alan M., GAHEGAN Mark, PIKE William, BREWER Isaac, CAI Guoray, LENGERICH Eugene et HARDISTY Frank. Geovisualization for Knowledge Construction and Decision Support. *IEEE, Computer graphics and application*, 2004, Vol. 24, n° 1, 13-17 p.

(MacEachren et Kraak, 1997) MACEACHREN Alan et KRAAK Menno-Jan. Exploratory cartographic visualization: advancing the agenda. *Computers & Geosciences*, 1997, Vol. 23, n° 4, 335-343 p.

(MacEachren et Kraak, 2001) MACEACHREN Alan M et KRAAK Menno-Jan. Research challenges in geovisualization, *Cartography and Geographic. Information Systems*, 2001, Vol 6, n°1, 3-12 p.

(McMaster et Shea, 1992) McMASTER Robert B. et SHEA Stuart K. Generalization in Digital Cartography. Washington : Association of American Geographers, 1992, 134 p

(Malinowsky et Zimányi. 2004) MALINOWSKI Elzbieta et ZIMÁNYI Esteban. Representing spatiality in a conceptual multidimensional model. In : PFOSER Dieter, CRUZ Isabel F. et RONTHALER Marc. 12th ACM International Workshop on Geographic Information Systems, 12-13 Novembre, 2004, Washington, DC, USA. New York, USA : ACM Press, 2004, 12-22 p.

(Malinowski et Zimányi, 2004a) MALINOWSKI Elzbieta et ZIMÁNYI Esteban. OLAP Hierarchies: A Conceptual Perspective. In : PERSSON Anne et STIRNA Janis. 16th International Conference Advanced Information Systems Engineering, 7-11 Juin, 2004, Riga, Latvia. Berlin Heidelberg : Springer, 2004, 477-491 p. (Lecture Notes in Computer Science 3084)

(Malinowsky et Zimányi 2005) MALINOWSKI Elzbieta et ZIMÁNYI Esteban. Spatial Hierarchies and Topological Relationships in SpatialMultiDimER model. In : JACKSON Mike, NELSON David et STIRK Sue. 22nd British National Conference on Databases, 5-7 Juillet 2005, Sunderland, UK. Berlin Heidelberg : Springer, 2005, 17-28 p. (Lecture Notes in Computer Science 3567)

(Malinowski, 2006) MALINOWSKI Elzbieta. Designing traditional, spatial, and temporal data warehouses: Concepts and methodological framework. Brussels : Faculté des Sciences appliquées, 2006, 103 p.

(Malinowski et Zimányi, 2007) MALINOWSKI Elzbieta et ZIMÁNYI Esteban. Logical Representation of a Conceptual Model for Spatial Data Warehouses. *Geoinformatica*, 2007

(Malinowski et Zimányi, 2007a) MALINOWSKI Elzbieta et ZIMÁNYI Esteban. Spatial Data Warehouses: Some Solutions and Unresolved Problems. In : 3th IEEE

International Workshop on Databases for Next-Generation Researchers, SWOD, 15 Avril 2007, Istanbul, Turquie.

(MapInfoLI, 2007) MAPINFO. Business Intelligence. [en ligne]. Disponible sur : <[http://www.mapinfo.com/location/integration?txtLeftNav=96269fe27443df00vcmprd00000034021dc7\\_\\_\\_\\_&txtTopNav=4a4a2545d8a37f00dev-vcml00001a031dc7\\_\\_\\_\\_](http://www.mapinfo.com/location/integration?txtLeftNav=96269fe27443df00vcmprd00000034021dc7____&txtTopNav=4a4a2545d8a37f00dev-vcml00001a031dc7____)> (consulté le 1.09.2007).

(MapInfoMJ, 2007) MAPINFO. MapXtreme Java. [en ligne]. Disponible sur : <<http://extranet.mapinfo.com/products/Overview.cfm?productid=1162>> (consulté le 1.09.2007).

(Marchand et al. 2003) MARCHAND Pierre, BRISEBOIS Alexandre, BÉDARD, Yvan et EDWARDS Geoffrey. Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis. Elsevier : Journal of the International Society of Photogrammetry and Remote Sensing, 2003, Vol. 59, n° 1, 6-20 p.

(Matias et Moura-Pires, 2005) MATIAS Rosa et MOURA-PIRES Joao. Spatial On-Line Analytical Processing: A Tool to Analyze the Emission of Pollutants in Industrial Installations. In : BENTO Carlos, CARDOSO Amílcar et DIAS Gaël. 5th International Workshop on Extraction of Knowledge from Databases and Warehouses, 5-8 Decembre, Covhila, Portugal.

(Matias et Moura-Pires, 2007) MATIAS Rosa et MOURA-PIRES Joao. Revisiting the OLAP Interaction to Cope with Spatial Data and Spatial Analysis. In : CARDOSO Jorge, CORDEIRO Jose et FILIPE Joaquim. 9th International Conference on Enterprise Information Systems, 12-16 Juin, 2007, Funchal, Madeira, Portugal. Portugal :INSTICC, 2007.

(MDX, 2007) MICROSOFT. [en ligne]. Disponible sur : <<http://msdn2.microsoft.com/en-us/library/ms145595.aspx>> (consulté le 1.09.2007).

(Mendelzon et Vaisman, 2003) MENDELZON Alberto O. et VAISMAN Alejandro A.. Time in Multidimensional Databases. In : RAFANELLI Maurizio. Multidimensional databases: problems and solutions. Hershey, PA, USA : IGI Publishing, 2003, 166-199 p.

(Mitchell, 2005) MITCHELL Andy. The ESRI Guide to GIS Analysis: Volume 2: Spatial Measurements and Statistics. Redlands, Californie : ESRI Press, 2005, 252 p.

(Miquel et al., 2002) MIQUEL Maryvonne, BEDARD Yvan et BRISEBOIS Alexandre. Conception d'entrepôts de données géospatiales à partir de sources hétérogènes Exemple d'application en foresterie. Ingénierie des Systèmes d'Information, 2002, Vol 7, n° 3, 89-111 p.

(Mohraz, 2000) MOHRAZ Karim. Geographical analysis in SAP business information warehouse. In : LI Ki-Joune, MAKKI Kia, PISSINOU Niki et RAVADA Siva. 8th ACM Symposium on Advances in Geographic Information Systems, 10-11 Novembre, 2000, Washington D.C., USA. Washington D.C., USA : ACM Press, 2000, 191-193 p.

(Mondrian, 2007) PENTAHO. [en ligne]. Disponible sur : <http://mondrian.pentaho.org/> (consulté le 1.09.2007).

(Nguyen et al., 2000) NGUYEN Thanh Binh, TJOA Min et WAGNER Roland. An object oriented multidimensional data model for olap. In : LU Hongjun et ZHOU Aoying. International Conference on Web-Age Information Management, 21-23 Juin, 2000, Shanghai, Chine. Berlin Heidelberg : Springer, 2000, 69-82 p. (Lecture Notes in Computer Science 1846)

(Newell, 1990) NEWELL Allen. Unified Theories of Cognition. Cambridge (Mass.) : Harvard University, 1990, 576 p.

(Neun et Steiniger, 2005) NEUN Moritz et STEINIGER Stefan. Modelling cartographic relations for categorical maps. In 12nd International Cartographic Association Conference, 2005, Lacoruna, Espagne.

(OpenGIS, 1999) OPEN GIS CONSORTIUM. OpenGIS simple features specification for SQL. Technical, 1999

(OracleBI, 2007) Oracle. [en ligne]. Disponible sur : <http://www.oracle.com/technology/products/bi/enterprise-edition.html> (consulté le 1.09.2007).

(OracleSpatial, 2007) ORACLE. Oracle® Spatial User's Guide and Reference Release 9.2 [en ligne]. Disponible sur : [http://download.oracle.com/docs/cd/B10501\\_01/appdev.920/a96630/toc.htm](http://download.oracle.com/docs/cd/B10501_01/appdev.920/a96630/toc.htm) (consulté le 1.09.2007).

(Oracle9i, 2007) ORACLE. Oracle9i® Database [en ligne]. Disponible sur : <http://www.oracle.com/technology/products/oracle9i/index.html> (consulté le 1.09.2007).

(Oracle10g, 2007) ORACLE. Oracle9i® Database [en ligne]. Disponible sur : [http://download.oracle.com/docs/html/B10827\\_01/geor\\_intro.htm](http://download.oracle.com/docs/html/B10827_01/geor_intro.htm) (consulté le 1.09.2007).

(Papadias et al., 2002) PAPADIAS Dimitris, TAO Yufei, KALNIS Panagiotis et ZHANG Jun. Indexing spatio-temporal data warehouses. In : 18th International Conference on Data Engineering, 26 Fevrier-1 Mars, 2002, San Jose, CA, USA. Los Alamitos, CA, USA: IEEE Computer Society, 2002, 166-175 p.

(Pedersen, 2000) PEDERSEN Torben Bach. Aspects of data modeling and query processing for complex multidimensional data. Thèse. Aalborg : University of Aalborg, 2000, 180 p.

(Pedersen, et al., 2001) PEDERSEN Torben Bach, JENSEN Christian S. et DYRESON Curtis E. A foundation for capturing and querying complex multidimensional data. Information Systems, 2001, Vol. 26, n° 5, 383-423 p.

(Pedersen et al., 2002) PEDERSEN Dennis, RIIS Karsten et PEDERSEN Torben Bach. A Powerful and SQL-Compatible Data Model and Query Language for OLAP. In : Xiaofang Zhou. 13th Australasian Database Conference, 26-30 Janvier 2002, Melbourne, Victoria. Australian Computer Society, 2002, Vol. 2

(Pedersen et Tryfona, 2001) PEDERSEN Torben Bach et TRYFONA Nectaria. Pre-aggregation in Spatial DataWarehouses. In : Christian S. Jensen, Markus Schneider, Bernhard Seeger, Vassilis J. Tsotras. 7th International Symposium on Spatial and Temporal Databases, 12-15 Juillet, 2001, Redondo Beach, CA, USA. New York : Springer, 2001, 460-478 p. (Lecture Notes in Computer Science 2121)

(Pendse, 2007) PENDSE Nigel. The OLAP Survey 6 [en ligne]. Disponible sur : <<http://www.survey.com/olap/>> (consulté le 1.09.2007).

(Pentaho) PENTAHO. [en ligne]. Disponible sur : <<http://www.pentaho.org>> (consulté le 1.09.2007).

(Persson et al., 2005) PERSSON Donata, GARTNER Georg et BUCHROITHNER Manfred. Towards Typology of Interactivity Functions for Visual Map Exploration. In : STEFANAKIS Emmanuel, PETERSON Michael P., ARMENAKIS Costas et DELIS Vasilis . 1st International Workshop on Geographic Hypermedia, 4-5 Avril, 2005, Denver, Colorado. Berlin Heidelberg : Springer, 2005, 275-292 p. (Lecture Notes in Geoinformation and Cartography)

(Pestana et al., 2005) PESTANA Gabriel et SILVA Miguel M. et BÉDARD, Yvan. Spatial OLAP Modeling: An Overview Based on Spatial Objects Changing over Time. In : IEEE 3th International Conference on Computational Cybernetics, 13-16 Avril, 2005, Ile Maurice, 149-154 p.

(Rafanelli, 2003) RAFANELLI Maurizio. Operators for Multidimensional Aggregate Data. In : Multidimensional databases: problems and solutions. Hershey, PA, USA : IGI Publishing, 2003, 116-165 p.

(Pourabbas, 2003) POURABBAS Elaheh. Cooperation with Geographic Databases. In : In : Multidimensional databases: problems and solutions. Hershey, PA, USA : IGI Publishing, 2003, 393-432 p.

(Pourabbas et Rafanelli, 2002) POURABBAS Elaheh et RAFANELLI Maurizio. A Pictorial Query Language for Querying Geographic Databases using Positional and OLAP Operators. SIGMOD Record, 2002, Vol. 31, n° 2, 22-27 p.

(Regnauld et McMaster, 2007) REGNAULD Nicolas et MCMASTER Robert B. A synoptic View of Generalization Operators. In : MACKANESS William A., RUAS Anne, SARJAKOSKI Tina Generalisation of Geographic Information Cartographic Modelling and Applications. Amsterdam : Elsevier Science, 2007, 386 p.



(Rivest, 2000) RIVEST Sonia. Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG. Rapport de DEA Informatique, Laval : Université Laval, Canada, 2000, 84 p.

(Rigaux, et al., 2001) RIGAUX Philippe, SCHOLL Michel et VOISARD Agnès. Spatial databases with application to GIS. 2<sup>nd</sup> Ed. San Francisco, CA : Morgan Kaufmann Publishers Inc., 2001, 440 p.

(Rigaux et Scholl, 1995) RIGAUX Philippe et SCHOLL Michel. Multi-Scale Partitions: Application to Spatial and Statistical Databases. In : EGENHOFER Max J. et HERRING John R. 4th International Symposium Advances in Spatial Databases, Aout 6-9, 1995, Portland, Maine, USA. Berlin: Springer, 1995, 170-183 p. (Lecture Notes in Computer Science 951)

(Rivest et al., 2001) RIVEST Sonia, BÉDARD Yvan et MARCHAND Pierre. Towards better support for spatial decision-making: defining the characteristics of Spatial On-Line Analytical Processing, *Geomatica*, Journal of the Canadian Institute of Geomatics, 2001, Vol. 55, n° 4, 539-555 p.

(Rivest et al. 2003) RIVEST Sonia, BÉDARD Yvan, PROULX Marie-Josée et NADEAUM Martin. SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis. International Society for Photogrammetry and Remote Sensing joint Workshop on Spatial, Temporal and Multi-Dimensional Data Modelling and Analysis, 2-3 Octobre, 2003, Quebec, Canada.

(Rivest et al. 2005) RIVEST Sonia, BÉDARD Yvan, PROULX Marie-Josée, NADEAUM Martin, HUBERT Frederic et PASTOR Julien. SOLAP: Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data. Elsevier : Journal of International Society for Photogrammetry and Remote Sensing, 2005, Vol. 60, n° 1, 17-33 p.

(Rizzi, et al., 2006) RIZZI Stefano, ABELLÓ Alberto, LECHTENBÖRGER Jens et TRUJILLO Juan. Research in data warehouse modeling and design: dead or alive ?. In : SONG Il-Yeol et VASSILIADIS Panos. 9th International Workshop on Data Warehousing and OLAP, 10 Novembre 2006, Arlington, Virginia, USA. New York, USA : ACM Press, 2006, 3- 10 p.

(Pourabbas, 2003) POURABBAS Elaheh. Cooperation with Geographic Databases. In : RAFANELLI Maurizio. Multidimensional databases: problems and solutions. Hershey, PA, USA : IGI Publishing, 2003, 393-432 p.

(Sampaio et al., 2006) SAMPAIO Marcus Costa, SOUSA Andre Gomes et BAPTISTA Cláudio. Towards a logical multidimensional model for spatial data warehousing and OLAP. In : SONG Il-Yeol et VASSILIADIS Panos. 9th ACM International Workshop on Data Warehousing and OLAP, 10 Novembre, 2006, Arlington, Virginia, USA. New York, USA : ACM Press, 2006, 83-90 p.

(Schneider, 2003) SCHNEIDER Michel. Well-formed data warehouse structures. In : LENZ Hans-Joachim , VASSILIADIS Panos, JEUSFELD Manfred A. et STAUDT Martin. 5th International Workshop of Design and Management of Data Warehouses, 8 Septembre, 2003, Berlin, Allemagne.

(Scotch, et Parmanto, 2005) SCOTCH Matthew et PARMANTO Bambang. SOVAT: Spatial OLAP Visualization and Analysis Tool. In : 38nd Hawaii International Conference on System Sciences, 3-6 Janvier, 2005, Hawaii. DC, USA : IEEE Computer Society, 2005, 142-149 p.

(Scotch, et Parmanto, 2006) SCOTCH Matthew et PARMANTO Bambang. Development of SOVAT: a numerical-spatial decision support system for community health assessment research. International Journal of Medical Informatics, 2006, Vol. 34, n° 10, 771-784 p.

(Serra, 1969) SERRA Jean. Introduction à la Morphologie Mathématique. Cahiers du Centre de Morphologie Mathématique, Fascicule n° 3, Ecole des Mines de Paris 1969, 160 p.

(Shekar et al., 2001) SHEKHAR Shashi, LU Chang-Tien, TAN X., CHAWLA Sanjay et VATSAVAI R. Map Cube: A Visualization Tool for Spatial Data Warehouses. In : MILLER Harvey J. et HAN Jiawei. Geographic Data Mining and Knowledge Discovery. Londres : Taylor & Francis, 2001, 74-109, p.

(Shneiderman et Plaisant, 2004) SHNEIDERMAN Ben et PLAISANT Catherine. Designing the User Interface: Strategies for Effective Human-Computer Interaction. 4<sup>nd</sup> Ed. Addison-Wesley, 2004, 672 p.

(Silva, et al., 2006) SILVA Joel, TIMES Valéria C. et SALGADO Ana C. An open source and web based framework for geographic and multidimensional processing. In : HADDAD Hisham. ACM Symposium on Applied Computing, 23-27 Avril, 2006, Dijon, France. New York : ACM Press, 2006, 63-67 p.

(Soumahoro, 2005) SOUMAHORO Inza. Intégration des concepts de la graphique dans la navigation SOLAP. Rapport de DEA. DEA Informatique, Lyon, France: INSA de Lyon, 2005, 29 p.

(Spaccapietra, et al., 2007) SPACCAPIETRA Stefano, PARENT Christine et ZIMANYI Esteban. Spatio-Temporal and Multi-Representation Modeling: A Contribution to Active Conceptual Modeling. In : Active Conceptual Modeling for Learning. Berlin Heidelberg : Springer, 2007 (Lecture Notes in Computer Science 4512)

(Stefanovic et al. 2000) STEFANOVIC Nebojsa, HAN Jiawei et KOPERSKI Krzysztof. Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. IEEE Transactions on Knowledge and Data Engineering, 2000, Vol. 12, n° 6, 938-958 p.

(Stell et Worboys, 1998) STELL John et Worboys Michael. Stratified Map Spaces: A formal basis for multi-resolution spatial databases. In :8th International Symposium on Spatial Data Handling, 11-15 Juillet 1998, Vancouver, Canada, 180-189 p.

(Stolte et al., 2002), STOLTE Chris, TANG Diane et HANRAHAN Pat. Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. IEEE Transactions on Visualization and Computer Graphics, 2002, Vol. 8, n° 1, 52-65 p.

(Stolte et al., 2003) STOLTE Chris, TANG Diane et HANRAHAN Pat. Multiscale visualization using data cubes. IEEE Transaction on Visualization and Computer Graphics, 2003, Vol. 9, n° 2, 176-187 p.

(Tableau, 2007) TABLEAU SOFTWARE. [en ligne]. Disponible sur : <<http://www.tableausoftware.com/>> (consulté le 1.09.2007).

(Tang et al., 2003) TANG Diane, STOLTE Chris et BOSCH Robert. Design Choices when Architecting Visualizations. In : 9th IEEE Symposium on Information Visualization, 20-21 Octobre 2003, Seattle, WA, USA. Piscataway, N.J : IEEE Computer Society, 2003

(Tchounikine et al., 2005) TCHOUNIKINE Anne, MIQUEL Maryvonne, LAURINI Robert, AHMED Taher, BIMONTE Sandro et BAILLOT Virginie. Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes. Revue des Nouvelles Technologies de l'Information - Entrepôts de données et l'Analyse en ligne. France : Cépaduès-Éditions, 2005, 21-33 p.

(Teste, 2001) TESTE Olivier. Towards Conceptual Multidimensional Design in Decision Support Systems. In : CAPLINSKAS Andres et EDER Johann. 5th East-European Conference on Advances in Databases and Information Systems, September 25-28, 2001, Vilnius, Lithuanie.. Research Communications, 2001, Vol. 1, 77-88 p.

(Thomsen, 1997) THOMSEN Erik. OLAP Solutions: Building Multidimensional Information Systems. New York,USA : Wiley, 1997, 608 p.

(Thomsen et Pedersen, 2005) THOMSEN Christian et PEDERSEN Torben Bach. A Survey of Open Source Tools for Business Intelligence. In : TJOA Min et TRUJILLO Juan. 7th International Conference on Data Warehousing and Knowledge Discovery, 22-26 Aout, 2005, Copenhagen, Danemark. Berlin Heidelberg : Springer, 2005, 74-84 p. (Lecture Notes in Computer Science 3589)

(Timpf, 1998) TIMPF Sabine. Hierarchical structures in map series. Thèse. Vienne : Technical University Vienna, 1998, 124 p.

(Timpf et Frank, 1995) TIMPF Sabine et FRANK Andrew U. A Multi-Scale DAG for Cartographic Objects. In : 12th International Symposium on Computer-Assisted Cartography, 27 Février-1 Mars, 1995, Charlotte, North Caroline, USA. ACM/ASPRS, 1995, Vol. 4, 157-163 p.

(Tomlin, 1990) TOMLIN C. Dana. Geographic Information Systems and Cartographic Modeling. New Jersey : Editeur : Prentice Hall, Englewood Cliff, 1990, 572 p.

(Torlone, 2003) TORLONE Riccardo. Conceptual multidimensional models. In : RAFANELLI Maurizio. Multidimensional databases: problems and solutions. Hershey, PA, USA : IGI Publishing, 2003, 69-90 p.

(Trujillo et al., 2000) TRUJILLO Juan, PALOMAR Manuel, GÓMEZ Jaime. An Object Oriented Approach to Multidimensional Databases & OLAP Operations. International Journal of Computer and Information Science, 2000, Vol. 1, n° 2, 75-85 p.

(Trujillo et al., 2001) TRUJILLO Juan, PALOMAR Manuel, GOMEZ Jaime, SONG Il-Yeol. Designing Data Warehouses with OO Conceptual Models. Computer, 2001, Vol. 34, n° 12, 66-75 p.

(Van den Worm, 2001) VAN DEN WORM Jeroen. Web map design in practice. In : KRAAK Jan-Menno et BROWN Allan. Web Cartography. Londres, UK : CRC, 2001, 213 p.

(Vangenot, 2001) VANGENOT Christelle. Supporting decision-making with alternative data representations. Journal of Geographic Information and Decision Analysis, 2001, Vol. 5, n° 2, 66-82 p.

(Vassiliadis, 1998) VASSILIADIS Panos. Modeling Multidimensional Databases, Cubes and Cube Operations. In : RAFANELLI Maurizio et JARKE Matthias. 10th International Conference on Scientific and Statistical Database Management, 01-03 Juillet, 1998, Capri, Italie. Los Alamitos, Calif. : IEEE Computer Society 1998, 53-62 p.

(Visualizer) COGNOS. Cognos Visualizer. [en ligne]. Disponible sur : <[http://www.cognos.com/products/business\\_intelligence/visualization\\_dashboards/index.html](http://www.cognos.com/products/business_intelligence/visualization_dashboards/index.html)> (consulté le 1.09.2007).

(Voisard et David, 2002) VOISARD Agnès et DAVID Benoit. A Database Perspective on Geospatial Data Modeling. In IEEE IEEE Transactions on Knowledge and Data Engineering, 2002, Vol. 14, n° 2, 26-243 p.

(Voss et al., 2004) VOSS Angi, HERNANDEZ Vera, VOSS Hans et SCHEIDER Simon. Interactive Visual Exploration of Multidimensional Data: Requirements for CommonGIS with OLAP. In : 15th International Workshop on Database and Expert Systems Applications, 30 Aout - 3 Septembre 2004, Saragosse, Espagne. Los Alamitos, Californie, USA : IEEE Computer Society, 2004, 883- 887 p.

(Weibel et Dutton, 2001) WEIBEL Robert et DUTTON Geoffrey. Generalizing Spatial Data and Dealing with Multiple Representations. In : LONGLEY Paul, GOODCHILD Michael, MAGUIRE David et RHIND David. Geographic Information Systems and Science. New York : John Wiley & Sons, 2001, 125-155 p.

(Winter, 1998) WINTER Richard. Databases: back in the OLAP game. *Intelligent Enterprise Magazine*, 1998, Vol. 1, 60–64 p.

(Worboys, 1995) WORBOYS Michael. *GIS: A Computing Perspective*. Londres : Taylor & Francis Group, 1995, 376 p.

(Zloof, 1975) ZLOOF Moshé M. Query-by-Example: the Invocation and Definition of Tables and Forms. In : KERR Douglas S.. *International Conference on Very Large Data Bases*, 22-24 Septembre, 1975, Framingham, Massachusetts, USA. New York : ACM, 1975, 1-24 p.