

République Algérienne Démocratique et Populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Amar Telidji

Laghouat

Faculté des Sciences

Département de Mathématiques et Informatique

Ecole Doctorale STIC

(Science et Technologie de l'Information et de la Communication)



Mémoire

En vue de l'obtention du Diplôme de :

Magistère en Informatique

Option : Informatique Répartie et Mobile (IRM)

Présenté par :

MAÏCHA MOHAMMED EL HABIB

Thème ;

Vers l'intégration de l'information géographique dans les
entrepôts de données;

Cas de l'historique du Bâti dans la ville de Laghouat

Soutenu publiquement devant le jury composé de :

- M. Yagoubi Med. Bachir	(PROF, université Amar Telidji de Laghouat)	Président
- M. Bensaad Lahcen	(MCB, université Amar Telidji de Laghouat)	Examineur
- M. Khorsi Ahmed	(Dr. Université Imam Saoud KSA)	Examineur
- Mme. Cherroun Hadda	(MCA, université Amar Telidji de Laghouat)	Encadreur
- Mme. Kerrouche Badra	(MAA, université Amar Telidji de Laghouat)	Co-encadreur

ملخص

ما زالت الجهود مبذولة من أجل تقديم أنظمة للمساعدة على اتخاذ القرارات على شبكة الأنترنت. غالبا ما تركز هذه النظم على أنظمة المعلومات الجغرافية ويكون ذلك لإنتاج المعلومات الجغرافية الداعمة لاتخاذ القرار. ومع ذلك، تبقى تكاليف التطوير مرتفعة جدا مقارنة بالأداء المقدم. ويرجع ذلك الى الافتقار للمعايير وأساليب النمذجة وكذا تعددية الاختصاصات اللازم توفرها لإعداد مثل هذه الأنظمة.

من خلال هذه المذكرة، نهدف لاستكشاف هذا المجال واكتساب الخبرات وكذا الدراية اللازمة لتصميم وتطوير نظام مساعد على اتخاذ القرار يستند الى مستودع بيانات جغرافية عبر الأنترنت. تصب اهتماماتنا خصوصا على ال مسائل المتعلقة بدمج المعلومة الجغرافية في عنصرها الزمني.

في هذا السياق، قمنا بتطوير منهجية للتصميم على أساس نمذجة البيانات متعددة الأبعاد، وكذا نظام مبني على مستودع بيانات جغرافية على شبكة الأنترنت تم انشاؤه باستخدام أدوات مفتوحة المصدر.

لإثبات جاهزية وفاعلية هذا النظام، اقترحنا دراسة التطور الزمني للعمران في مدينة الأغواط. الأداء والنتائج المحصل عليها واعدة.

الكلمات الرئيسية: معلومات جغرافية – مستودع بيانات جغرافية – نظام خرائط على شبكة الأنترنت – نظام المعلومات الجغرافية – أنظمة دعم اتخاذ القرارات

Résumé

D'importants efforts sont déployés en vue de mettre en place des systèmes d'aide à la décision sur Internet. Souvent ces systèmes reposent sur les systèmes d'information géographique (SIG) pour produire l'information géodécisionnelle. Cependant, leurs coûts de développement sont trop élevés à comparer avec les performances offertes. Ceci est dû aux manques de standards et méthodes de modélisation, ainsi qu'à la pluridisciplinarité exigée pour produire de tels systèmes.

Dans ce travail, on vise à explorer ce domaine et à acquérir des savoirs-faire liés à la conception et réalisation d'un système d'aide à la décision basé sur un entrepôt de données géographique via le Web (Web-EDG). Nous nous intéressons spécialement aux problèmes liés à l'intégration de l'information géographique dans sa composante temporelle.

Dans ce contexte, nous avons développé une méthode de conception se basant sur la modélisation des données multidimensionnelles. Le système Web-EDG est développé à l'aide d'outils Open Source.

Afin de prouver les performances de notre système, nous l'avons projeté sur l'étude de l'évolution du bâti de la ville de Laghouat. Les performances obtenus sont prometteuses.

Mots-clés : Information géographique, Entrepôt de données géographique, OLAP, Géodécisionnelle, Web-mapping, SIG.

Abstract

Major efforts are deployed in order to set up the decision support systems on the Internet. Often these systems are based on geographic information systems (GIS) to produce geodecisional information. However, development costs are too high compared with the performance offered. This is due to lack of standards and modeling methods, and multidisciplinary required to produce such systems.

In this work, we aim to explore this field and to acquire know-how related to the design and realization of a decision support system based on a geographical data warehouse via the Web (Web-EDG). We are especially interested in issues related to the integration of geographic information in its temporal component.

In this context, we have developed a design approach based on the modeling of multi-dimensional data. Web-EDG system is developed using Open Source tools.

In order to prove the performances of our system, we have projected it on the study of the evolution of Building for the town of Laghouat. The performances obtained are promising.

Keywords : Geographic Information, Geographic Data Warehouse, OLAP, BI, web-mapping, GIS.

Remerciements

Ce travail n'aurait jamais vu le jour sans la volonté d'**ALLAH** qui m'a offert santé, force, courage et volonté jusqu'au dernier moment. Je vous remercie **ALLAH** pour ça et pour tout le reste.

Je m'adresse tous mes remerciements et ma sincère reconnaissance à :

- *Mme. CHERROUN*, non seulement pour avoir accepté de m'encadrer et ainsi me faire profiter de ses connaissances, mais aussi pour sa patience et pour la totale confiance qu'elle m'a accordée,
- *Mme. KERROUCHE* qui a bien prodiguer ses précieux conseils et de son aide tout au long de ce travail et qui a eu la gentillesse de lire et corriger ce travail,
- Enfin, Messieurs le président et les membres du *jury*, d'avoir bien voulu me faire l'honneur de juger mon travail.

Aussi, j'adresse mes plus sincères remerciements à ma famille spécialement ma mère *ZOHRA*, tous mes proches et amis, qui m'ont toujours soutenue et encouragée au cours de la réalisation de ce mémoire. *Merci à tous et à toutes.*

Table des matières

Liste des tableaux	v
Table des figures	vii
Introduction générale	viii
1 L’information Géographique et les Entrepôts de données Géographique	1
1.1 Information géographique	2
1.2 Les systèmes d’information géographique	2
1.2.1 Concepts & Définitions	3
1.2.1.1 Raster	3
1.2.1.2 Vector	4
1.2.1.3 Couches	4
1.2.2 Les SIG et le Web(Web-mapping)	5
1.2.2.1 Fonctionnement du Web-mapping	5
1.2.3 Outils et Logiciels SIG	6
1.2.3.1 Logiciels libres	6
1.2.3.2 Logiciels propriétaires	6
1.3 Entrepôts de données & Analyse en ligne	7
1.3.1 Dimensions et hiérarchies	8
1.3.2 Faits et mesures	8
1.3.3 Hypercube	9
1.3.4 Requêtes multidimensionnelles	9
1.3.5 Architecture des entrepôts de données	10
1.3.5.1 Entrepôt de données	11
1.3.5.2 Serveur OLAP	11
1.3.5.3 Client OLAP	13
1.4 Conclusion	14

2	Travaux Connexes	15
2.1	Les Travaux de Bédard	16
2.2	Les travaux de Brisebois 2002	17
2.3	Les travaux Pestana & al. 2005	18
2.4	Les travaux de Bimonte & al. 2006, 2007	18
2.5	Travaux de Bimonte 2008	19
2.6	Travaux de Selmoune & al. 2012	20
2.7	synthèse, comparaison et discussion	21
3	Conception	23
3.1	Contexte : La gestion socio-spatiale en milieu semi-aride ; cas de la wilaya de Laghouat	24
3.2	Entrepôt de données géographique dans le Web	25
3.2.1	Modélisation dimensionnelle des données	25
3.2.1.1	Le choix du processus de modélisation dimensionnelle	26
3.2.1.2	Constellation de Faits	28
3.2.2	Validation du Modèle	28
3.2.2.1	Niveau de granularité	29
3.2.2.2	Hétérogénéité	29
3.2.2.3	Choix des dimensions	31
3.2.2.4	Identification des faits	33
3.2.3	Intégration des données	33
3.3	Bilan	35
4	Étude expérimentale	37
4.1	Démarche d'implémentation du prototype Web-mapping à base d'EDG	38
4.1.1	Architecture de la solution Web-EDG	38
4.1.1.1	L'entrepôt de données géographique	38
4.1.1.2	Le Serveur GeoServer	39
4.1.1.3	Le client Web	40
4.1.2	Interface du système	40
4.1.2.1	La Vue générale	41
4.1.2.2	La Vue Comparative	42
4.1.2.3	L'analyse temporelle	42
4.1.3	Fonctionnalités	43
4.2	Test de Performances	44

4.2.1	Satisfaction des utilisateurs	44
4.2.2	Délais d'attente	45
4.3	Analyse comparative	47
4.4	Limites	48
4.5	Conclusion	49
	Conclusion générale	51
	Annexe A	55
.1	GetFeaturInfo()	56
.2	MapFrame	57
.3	Le TimeLine	58
.4	Le dictionnaire de données	59
	Bibliographie	61

Liste des tableaux

1.1	Principales Différences entre les systèmes OLPT et OLAP	7
1.2	Correspondance entre les opérateurs de BDS et OLAP	10
2.1	Revu sur l'ensemble de travaux antérieurs	21
3.1	Les domaines d'étude	26
4.1	Résumé de la configuration du GeoServer	44
4.2	Résumé des résultats de test en terme de tâches réelles effectuées	46

Table des figures

1.1	La composante spatiale des communes de Laghouat	3
1.2	L'information Vector et Raster	3
1.3	Structuration des informations géographiques en couches	5
1.4	Architecture d'un serveur cartographique sur internet	5
1.5	Hierarchie des communes a) Schéma b) Instance	8
1.6	Application multidimensionnelle a) Schéma b) Hypercube	9
1.7	Architecture à trois niveau d'un système d'entrepôt de données	11
1.8	a)Schéma en étoile b)Schéma en flocon c)Schéma en constellation	13
2.1	Évolution chronologique des modèles conceptuels de données pour les SIG	16
2.2	Architecture de GeWOlap	19
2.4	L'architecture de la solution ANRH	20
2.3	Processus de conception multidimensionnelle	20
3.1	Vue générale du système projeté	25
3.2	Modèles dimensionnels	27
3.3	Vue partielle du Modèle en constellation de faits envisagé	28
3.4	Vue simplifiée du Modèle en constellation de faits précédent	28
3.5	Schéma en étoile du modèle bâti	33
3.6	Instance du processus d'implémentation de GESSlagh	34
4.1	Architecture de GESSlagh	38
4.2	La page principale de l'interface GESSlagh	41
4.3	L'onglet : Vue générale	41
4.4	L'onglet : Vue générale - Info Tool	42
4.5	L'onglet : Vue générale - Contrôle de couches	42
4.6	L'onglet Vue Comparatif : Split	42
4.7	L'onglet TimeL : Analyse temporelle	43

4.8	Tests de performances sur l'onglet Vue simple	45
4.9	Tests de performances sur l'onglet Vue comparative	45
4.10	Tests de performance sur l'onglet TimeLine	45
4.11	Tests de performances de GESSlagh sur un réseau proche de la réalité - tâches en arrière plan	47
4.12	Tests de performances de GESSlagh sur un réseau proche de la réalité. .	47

Introduction Générale

Contexte

A l'ère des technologies de l'information, le volume de données que l'on stocke ne cesse de grandir. On organise et on classe ces données dans des différents types de structures particulières tel que les Bases des données, Fichiers, Tableaux, états statistiques,...etc. Vers la fin des années 90 et pendant la révolution de l'informatique décisionnelle, sont apparus les entrepôts de données augmentés par les outils OLAP (On Line Analytical Processing) permettant d'interroger et analyser des données structurées de façon aisée et conviviale [Ie07].

On appelle *Information géographique*, la représentation des objets ou d'un phénomène réel localisés dans l'espace. Cette information, comme nous allons la décrire dans le chapitre 1, est caractérisée par sa localisation, sa forme et par ses aspects sémantiques [PMA01].

Les premières cartes eurent plusieurs inconvénients : quantité d'informations limitée, documents figés dont la mise à jour est difficile, manipulation de cartes à échelles différentes malaisée. C'est pourquoi, on a essayé de limiter ces inconvénients en profitant de l'essor de l'informatique, Ce qui explique la naissance des Systèmes d'Information Géographique (SIG) [ASCR].

En 1989, la société française de Photogrammétrie et de télédétection définit le SIG comme étant "*un système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer, d'analyser et de combiner, d'élaborer et de présenter des informations localisées géographiquement (géoréférencées). L'ensemble des informations géographiques intégrées dans le SIG forment une base de données géographiques.*" [SFP89]. Un logiciel SIG est un système de gestion de bases de données (SGBD) localisées qui comprend une ou plusieurs couches géographiques qui peuvent entretenir des relations entre elles, être croisées et interrogées [DS04].

Parmi les plus importantes lacunes des SIG, on décèle leur incapacité à supporter les données provenant de plusieurs sources hétérogènes et relevées en différentes époques. Afin de pallier à cette lacune, deux solutions ont été proposées : les bases de données privées ou les entrepôts de données.

En effet, les bases de données opérationnelles sont utilisées par les services pour la production et la gestion de leurs données. C'est dans ces bases que sont stockées les données attributaires ou géoréférencées. Ces bases sont dites « privées » car elles n'existent que pour un domaine bien particulier et pour un ensemble de besoin défini.

L'entrepôt de données (data warehouse en anglais) est une base de données décisionnelle utilisée pour collecter, ordonner, archiver et stocker des informations provenant de bases de données opérationnelles [VBQ13]. leur objectif majeur est de partager l'ensemble des données utiles par les services, afin de faciliter l'utilisation de ces données par les utilisateurs du SIG et plus spécialement lorsque la diffusion de l'information géographique est faite via le réseau Internet. En effet, Le Web Mapping est un domaine en pleine expansion grâce au développement des solutions Open Source. [Mj01].

Motivations et Objectifs

Les modèles SOLAP existants considèrent une vision non continue de l'espace géographique [Bé97]. Nous suivons dans notre travail la même considération. Nous associons le terme *objet complexe* à une entité du monde réel (ex : un produit) décrit par un ensemble d'attributs descriptifs alphanumériques (nom, type, etc.) et pour représenter cet objet complexe on utilise le terme *objet géographique* (ex : une ville, un bâtiment) qui présente une géométrie (un attribut spatial). La Wilaya de Laghouat peut être un exemple d'un objet géographique qui est décrit par un nom (Laghouat), par une population (600.000) et par une géométrie représentée par un point sur la carte.

Il est difficile pour les utilisateurs finaux d'acquérir des données spatiales complètes et en temps réel à partir de plusieurs sources hétérogènes, pour cela un système consolidé accessible partout et par tous le monde doit être présent. Il nous a été confié de concevoir l'EDG de façon la plus générique possible pour pouvoir l'étendre à n'importe qu'elle domaine d'application. Le domaine d'étude choisi pour illustrer notre travail, est l'étude du bâti et son évolution chronologique.

Ainsi, notre recherche vise à faciliter le terrain :

a- Vers la mise en place d'un entrepôt de données géographique

Actuellement, il n'y a aucune chaîne de traitement et d'échange d'information qui regroupe l'ensemble des systèmes d'informations des différentes agences à travers la wilaya de Laghouat. Nos travaux de recherche rentrent dans le cadre d'un Projet National de Recherche PNR¹. Ce dernier vise à élaborer un outil d'aide à la décision pour la Wilaya de Laghouat, intégrant et synthétisant plusieurs types de données (Géo-économique, socio-spatiales... etc) provenant de plusieurs collaborateurs. La partie qui nous a été confiée est de concevoir un entrepôt de données géographique.

Le système pluridisciplinaire dédiée à tous les internautes demandeurs d'informations, suit une recherche multidimensionnelle pour produire les résultats, de façon similaire aux opérations d'analyse en ligne OLAP. Le traitement des données statiques d'un tel système et la structure conceptuelle d'un Entrepôt de données géographique ont exactement les mêmes composants : une mesure, une ou plusieurs dimensions avec ou pas de hiérarchisation.

b- Vers une mise en place d'une application SOLAP-SIG

Aujourd'hui, peu importe le type de demandeurs d'information à travers le Web, leur désir est le même ; Avoir des applications qui permettrait d'analyser et d'exploiter les données via des requêtes de façon interactive, conviviale, rapide et visuelle de préférence sur des cartes (i.e, applications SIG). En effet, les données socio-spatiales et statistiques concernant la wilaya de Laghouat obéissent à une recherche multidimensionnelle et donc les structurées en EDG permettrait une analyse multidimensionnelle consistant par exemple à explorer ces données d'un niveau détaillé à un niveau agrégé, et vis versa, grâce aux outils OLAP.

La consultation et l'étude des données via une interface type *Web-mapping* devient très facile et rapide. Par exemple, il est intéressant de pouvoir basculer entre plusieurs cartes si on veut comparer les données dans le temps ou dans leur mode d'analyse. Un

1. PNR : "Intégration de l'imagerie satellitale et des données socio-économiques pour la gestion socio-spatiale en milieu semi aride : cas de la Wilaya de Laghouat", projet financé par le MESRS Code :8/U03/7015

autre exemple et de pouvoir descendre jusqu'à un niveau de détail intéressant pour les utilisateurs.

Contributions

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

Dans ce contexte et pour atteindre les objectifs cités précédemment, nous avons apporté principalement 3 contributions

Une étude approfondie des concepts principaux du SIG, de l'information géographique, du web-mapping et de l'analyse spatio-temporelle, nous a mené à la définition d'un système d'analyse socio-territoriale déployé dans le Web basé sur un Entrepôt de données géographique. L'EDG reformule les concepts du ED pour prendre en compte la composante spatiale et sémantique de l'information géographique et la flexibilité de l'analyse spatiale. Ainsi qu'une nouvelle composante qu'on l'appelle "La composante temporelle" de l'information géographique.

1. Notre première contribution a consisté à rassembler les savoirs faire de l'informatique décisionnelle y compris les entrepôts de données géographiques, les concepts principaux du SIG et du Web-mapping.
2. Dans une deuxième contribution, nous avons étudié les travaux qui existent dans la littérature liées à la conception des EDG.
3. Dans une troisième contribution on a conçu une méthode générique de conception des entrepôts de données géographiques basées sur la modélisation multidimensionnelle, et nous avons appliqué cette méthode à notre cas de gestion du bâti et le suivi de son évolution.

Structure du document

Après cette introduction, nous structurons ce mémoire en cinq chapitres, présentant respectivement un état de l'art, un panorama des travaux antérieurs, notre proposition en deux chapitres, et une conclusion.

Le chapitre 1 décrit les principaux concepts de l'information géographique, des systèmes d'information géographiques ainsi les principaux concepts des entrepôts de données et de la modélisation multidimensionnelle, car nous nous appuyerons sur ces concepts pour la

formalisation de notre modèle. Le chapitre 1 présente également l'architecture classique des systèmes décisionnels et celle du Web-mapping.

Le chapitre 2 aborde une étude de certains travaux qui existent dans la littérature liés au SIG et ED, qui ont une relation avec notre travail. Cette analyse nous mène à une conclusion, que les Méthodes et les outils de conception des systèmes à base d'EDG dans le Web sont à leur état primitifs et ne s'appliquent qu'à des cas particuliers. En effet, très peu de travaux de généralisation et d'abstractions des concepts existent. Notre contribution est présentée dans le troisième et le quatrième chapitres.

Dans le chapitre 3 une description de l'architecture du système envisagé précède le processus de modélisation multidimensionnelle. Ainsi nous présentons dans ce chapitre notre modèle Web-EDG qui formalise les concepts de l'OLAP géographique et les aspects SIG dans le Web présentés au chapitre 1.

Dans le chapitre 4 nous montrons que les techniques classiques de visualisation cartographiques ne sont pas adéquates pour l'analyse temporelle des mesures géographiques. En utilisant une maquette d'interface, nous proposons un nouveau paradigme d'interaction et de visualisation : le SIGHout. A travers un cas d'étude concernant l'étude de l'évolution du Bâti, nous décrivons notre technique de visualisation et d'interaction et ses caractéristiques principales.

La conclusion générale, présente le bilan de notre travail et les perspectives de recherche que nous envisageons afin de compléter et améliorer cette étude.

Chapitre 1

L'information Géographique et les Entrepôts de données Géographique

De nombreuses années de collecte de données dans plusieurs domaines et différents systèmes ont généré des volumes importants d'informations hétérogènes. Informations qui non seulement doivent être stockées, mais aussi elles doivent être transformées pour les exploitées dans les processus d'analyse. Les entrepôts de données (DW) sont des systèmes qui permettent de stocker et d'explorer d'énormes quantités d'informations à partir de plusieurs sources de données.

Néanmoins, il ya des données qui ne sont pas utilisées à leur plein potentiel et une partie de leur richesse est tout simplement laissé de côté, ce sont les données spatiales. Cette limitation d'analyse est causée par la représentation alphanumérique typique qui est utilisé pour stocker et explorer ce type de données.

Un SIG (Système d'information géographique) est un Système informatique permettant, à partir de diverses sources, de rassembler et d'organiser, de gérer, d'analyser et de combiner, d'élaborer et de présenter ce type d'informations dite (localisées géographiquement), contribuant notamment à la gestion de l'espace¹. N'importe qui peut, un jour ou l'autre, être amené à utiliser un SIG qui a besoin d'analyser d'importants volumes de données géographiques dans différents domaines (urbanisme, environnement, climatologie, marketing, agriculture, santé, transport, etc.).

Dans ce chapitre, on va définir qu'est ce qu'une information géographique. Ensuite, nous passerons à décrire les fonctionnalités des systèmes d'information géographique et des entrepôts de données.

1. Société française de photogrammétrie et télédétection,1989

1.1 Information géographique

L'information géographique est la représentation d'un objet ou d'un phénomène réel ou imaginaire, présent, passé ou futur, localisé dans l'espace à un moment donné et quelles qu'en soient la dimension et l'échelle de représentation.

On distingue habituellement deux types d'information géographique :

- des informations de base ou de référence (ex. : référentiel à grande échelle) ;
- des informations thématiques concernant un domaine thématique particulier (environnement, transport, réseaux d'utilités, foncier, etc.) venant enrichir la description d'un espace ou d'un phénomène défini par des informations de base².

Composantes

Les trois composantes de l'information géographique sont :

le niveau sémantique .L'information relative à un objet décrit par sa nature, son aspect (ex. : le numéro d'une parcelle cadastrale, le nom d'une route, d'une rivière, d'une commune, etc.) ;

le niveau topologique .Les relations éventuelles avec d'autres objets ou phénomènes (ex. : la contiguïté entre deux communes, l'inclusion d'une parcelle dans une commune, l'adjacence entre les différents nœuds des tronçons constituant des parcelles cadastrales, etc.) ;

le niveau géométrique .Appelé aussi *la composante spatiale*, c'est la forme et la localisation de l'objet sur la surface terrestre, exprimés dans un système de coordonnées explicite (ex. : coordonnées géographiques polaires ou sphériques de type Longitude-Latitude ou coordonnées cartographiques issues d'une projection cartographique comme la projection Lambert). Dans la *Figure 1.1* est montrée la composante spatiale de l'information géographique représentant les communes de Laghouat. Chaque commune possède une forme et occupe une position précise par rapport aux autres communes.

1.2 Les systèmes d'information géographique

Comme il a été mentionné précédemment, Un système d'information géographique (SIG) est un système d'information permettant de créer, d'organiser et de présenter des

2. Numéro 119 de la revue *L'Information géographique*, (en 2012, via Persée)



FIGURE 1.1 – La composante spatiale des communes de Laghouat

données alphanumériques spatialement référencées, autrement dit géoréférencées, ainsi que de produire des plans et des cartes. Ses usages couvrent les activités géomatiques de traitement, de partage et de diffusion de l'information géographique[DS04].

Dans les années 1960, les cartes de l'Afrique de l'Est trop nombreuses pour permettre de localiser les meilleurs endroits pour créer de nouvelles implantations forestières font naître l'idée d'utiliser l'informatique pour traiter les données géographiques[KL99].

1.2.1 Concepts & Définitions

La représentation de l'information géographique se fait à travers deux types de modèles ou structures de données : les modèles vectoriel (*Vector*) et matriciel (*Raster*) [PMA01].

1.2.1.1 Raster

Le modèle matriciel 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 (*Figure 1.2*). Le choix dépend de la nature des données géographiques et de leur utilisation [Hel92]. Le modèle Raster 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*» [Dan90] est possible.[Bim07]

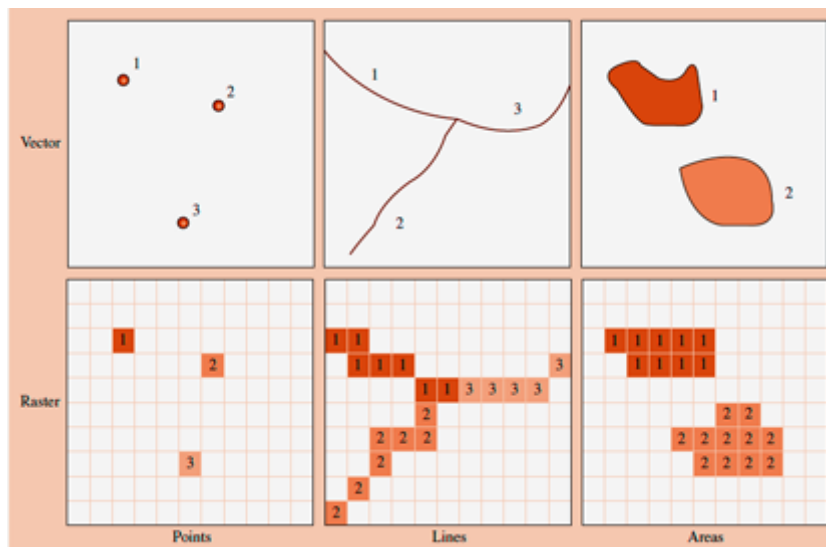


FIGURE 1.2 – L'information Vector et Raster

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. [Bim07]

1.2.1.2 Vector

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.) [PMDD01]. et grâce à leurs propriétés métriques (la surface, etc.). Plusieurs auteurs adoptent le modèle vectoriel pour définir l'information géographique. Philippe et al [PMA01], 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

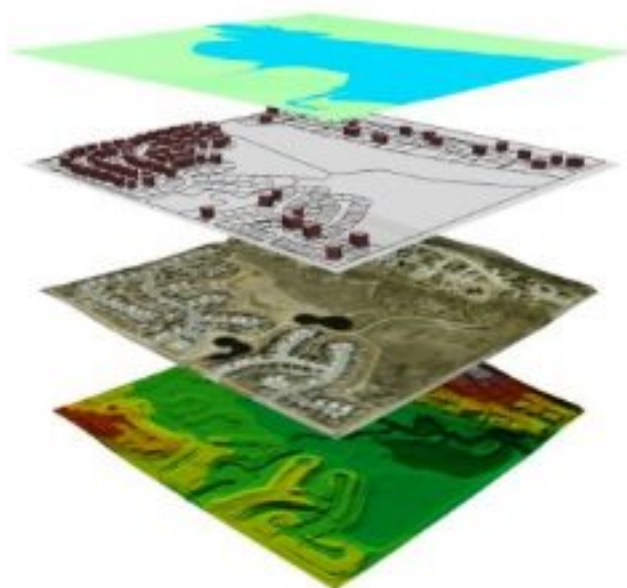


FIGURE 1.3 – Structuration des informations géographiques en couches

le nom, la population etc.) et la composante spatiale qui correspond à sa géométrie et ses relations topologiques. Worboys et Dokham [WD04] appelle un objet avec référence spatiale ou géo-objet, une entité avec des attributs spatiaux, temporels, graphiques et alphanumériques.[Bim07]

1.2.1.3 Couches

La métaphore utilisée dans les 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 1.3). Cette structuration rend ces informations superposables, compatibles, combinables et par conséquent analysables.[Bim07]

1.2.2 Les SIG et le Web(Web-mapping)

Aujourd'hui, les SIG ont révolutionné nombreux domaines, notamment avec l'avènement de l'Internet et du Web. En WebGIS, les technologies de l'internet sont connectés avec les SIG afin de tirer profit de leurs caractéristiques particulières, telles que la facilité d'utilisation, l'utilisation des données du SIG telles que le chargement, l'ajustement, la manipulation, l'analyse la production de l'information géographique. Alors que les SIG classiques dits autonomes n'avaient pas la capacité d'application sur les réseaux, le

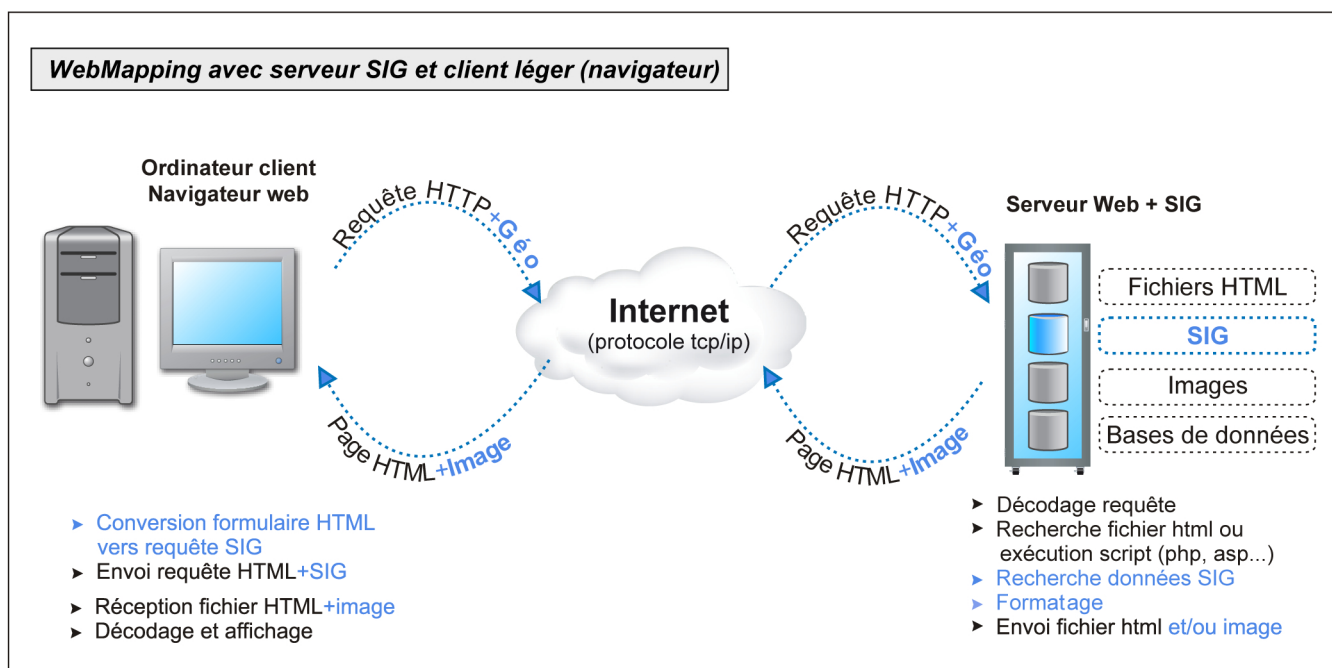


FIGURE 1.4 – Architecture d'un serveur cartographique sur internet

SIG internet permet de récupérer et d'analyser des données spatiales à travers le web [VVSN12].

1.2.2.1 Fonctionnement du Web-mapping

Le serveur cartographique est le guichet automatique auquel l'utilisateur fait appel pour afficher des cartes sur son poste informatique. Par le protocole de communication Internet, TCP/IP, des ordinateurs branchés en réseau peuvent échanger de l'information via un navigateur Web ou transférer des fichiers grâce au protocole FTP. L'architecture est de type client/serveur, c'est-à-dire qu'un ordinateur dit serveur répond aux requêtes d'une série d'ordinateurs dits clients (Figure 1.4) ³.

L'utilisateur, à partir de son terminal effectue des requêtes pour demander l'affichage d'une carte spécifique; le serveur cartographique interprète cette requête et renvoie la carte sous la forme d'une image matricielle (png, jpg,...) ou vectorielle (svg, swf,...).

Le moteur cartographique peut être contrôlé par des langages de script tels que PHP, javascript, Python ou Perl qui lui permettent de générer dynamiquement une carte en réponse à une requête préparée par une interface utilisateur. Le serveur cartographique peut chercher l'information nécessaire à la réalisation de la carte dans ses propres ressources,

3. Source : Notes de cours Serveur Cartographique et SIG interactifs en ligne M2 IASIG 2010

mais aussi sur des serveurs de données distants.

1.2.3 Outils et Logiciels SIG

Plusieurs logiciels SIG commerciales et gratuites (Open source) ont été développées dès leur première apparition en 1980s. nous proposons ici une liste non-exhaustive de logiciels utilisables dans ce domaine :

1.2.3.1 Logiciels libres

- **GeoServer** : Un serveur open source écrit en javascript qui permet aux utilisateurs de partager et modifier des données géospatiales
- **Quantum GIS** : logiciel de cartographie basé sur la bibliothèque Qt. Il est disponible sous Linux (KDE), Mac OS X, ou Windows. Entre autres choses, il permet la visualisation des couches de données comme des *shapefiles* ainsi que leur modification. Il permet notamment l'élaboration de fichiers destinés à être publiés sur *MapServer*. Il présente une ergonomie aboutie qui le rend très simple à utiliser
- **MapServer** : logiciel de publication de carte sur Internet. Il peut être utilisé pour réaliser des applications Web, mais également pour publier des services Web conformes aux recommandations de l'Open Geospatial Consortium (WMS, WFS, WCS) ;
- **OpenLayers** : permet d'intégrer des cartes glissantes dans des sites web
- **PostGIS** : extension pour la base de données PostgreSQL qui permet de faire des requêtes spatiales
- ...

1.2.3.2 Logiciels propriétaires

- **MapInfo**, **MapXtreme**, **MapMarker** de Pitney Bowes Software (PBS), leader mondial SIG et Géomarketing
- **ArcGIS** (ArcInfo, ArcView, ...) d'Esri - leader historique du marché du SIG
- **Oracle Spatial**, stockage d'informations géographiques et requêtes spatiales
- **IDRISI** (Clark Labs, de l'Université Clark) : logiciel SIG, principalement basé sur le mode raster (image)
- **Aigle Technologies** de Business Geografic : Générateur d'applications SIG full web, Solutions métiers packagées, Géo-Décisionnel avec GeoQlik pour QlikView, GeoBI pour Business Objects et Aigle Maps pour Google Maps

- ...

1.3 Entrepôts de données & Analyse en ligne

« Un entrepôt de données est une collection de données portant sur des sujets touchant une organisation, intégrée, variant dans le temps, et non-volatile pour supporter le processus de prise de décision » [W.H96]. 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).[Bim07]

Les différences principales entre les systèmes OLTP et OLAP résumées par Kelly [KEL97] sont soulignées dans le Tableau 1.1.

Pour que les entrepôts de données puisse faire des analyses multidimensionnelles de grandes quantités de données, ils reposent sur le paradigme multidimensionnel qui s'occupe de cette tâche grâce aux concepts de dimension, fait, mesure, hypercube et opérateurs OLAP.

1.3.1 Dimensions et hiérarchies

Dans l'analyse multidimensionnelle les axes sont représentées par les dimensions. ces derniers sont organisées en schémas hiérarchiques. Un schéma de hiérarchie, composé par plusieurs niveaux, représente différentes granularités ou degrés de précision de l'information[Bim07]. 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 le découpages de communes d'une Wilaya décrite par une hiérarchie avec les niveaux «Site» et «Commune» (Figure 1.5a). 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 communes est montré en (Figure 1.5b), où par exemple les sites «Bab D'zair» et «Zgag el Hedjaj»,

	OLTP	OLAP
Utilisation	<ul style="list-style-type: none"> - Gestion des transactions - Répétitif 	<ul style="list-style-type: none"> - Aide à la Décision - Ad hoc
Conception	Orientée applications	Orientée utilisateurs
Fréquence	Quotidienne	Sporadique
Données	Récentes, détaillées	Historiques, multidimensionnelles, agrégées
Source	BD unique	Plusieurs BD
Accès	Programmes précompilés	Outils d'Analyse
Nb lignes accédées	Dizaines	Milliers
Type d'utilisateur	Opérateurs	Décideurs
Nb utilisateurs	Milliers	Centaines
performance	Elevée	Basse
Dimension BD	Giga-octets	Téraoctets

TABLE 1.1 – Principales Différences entre les systèmes OLPT et OLAP
[KEL97]

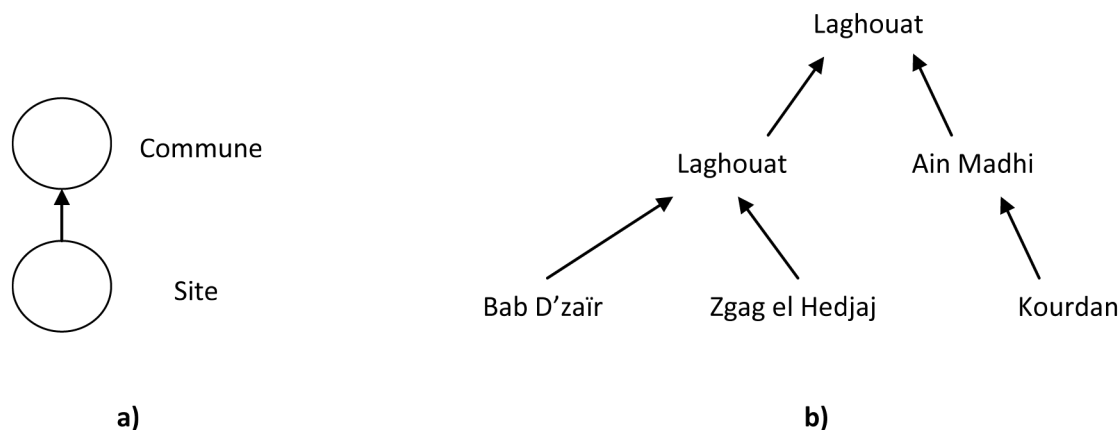


FIGURE 1.5 – Hiérarchie des communes a) Schéma b) Instance
[Bim07]

appartenant au niveau «Site», sont liés au membre «Laghouat» du niveau «Commune».

1.3.2 Faits et mesures

Au sein d'une organisation, la modélisation d'un ensemble d'événements lors d'un processus décisionnel donne naissance à un nouveau concept qui s'appelle *fait*. 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.[Bim07]

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 1.6a représente le schéma de cette application grâce au modèle conceptuel multidimensionnel présenté. 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 [Bim08].

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,

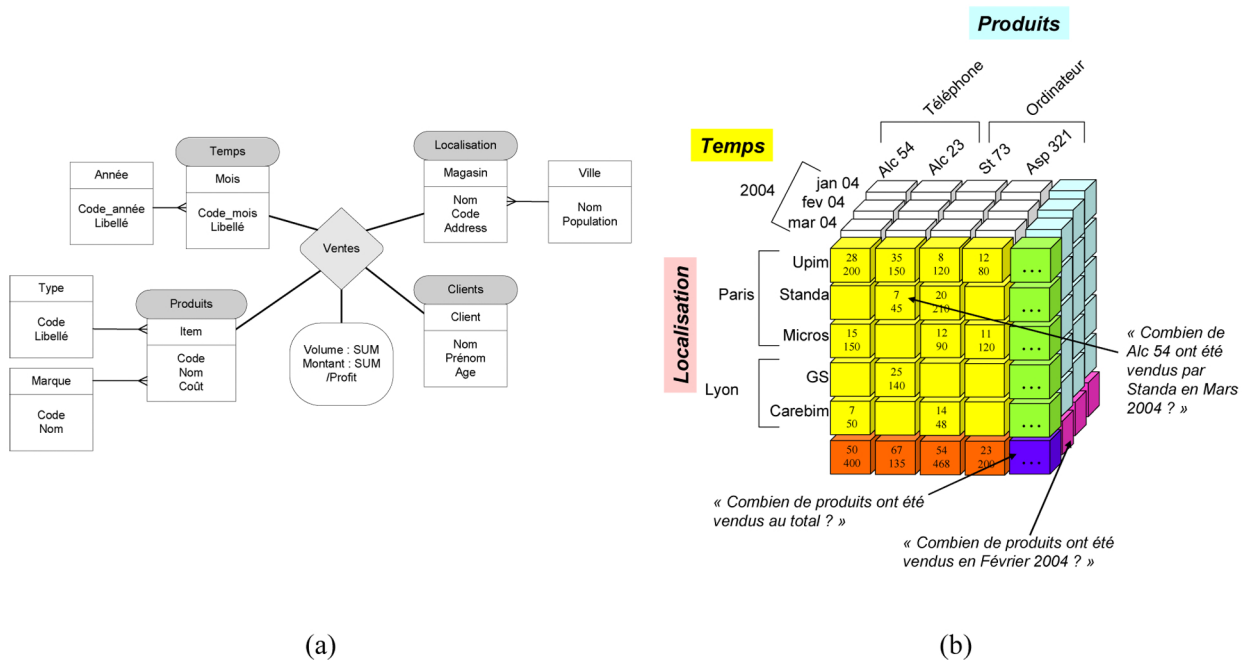


FIGURE 1.6 – Application multidimensionnelle a) Schéma b) Hypercube [Bim07]

en revanche le maximum est significatif[Bim07]. 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 [Ral96]. 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.

1.3.3 Hypercube

Un hypercube est l'instance d'un modèle conceptuel multidimensionnel est . Il 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 1.6a est montré en Figure 1.6b. Sur les axes du cube de base on trouve les membres des niveaux des dimensions et dans les cellules les valeurs des deux mesures[Bim07].

1.3.4 Requêtes multidimensionnelles

Les utilisateurs interagissent itérativement avec le modèle multidimensionnel pour formuler, modifier et valider leurs besoins. Les chemins d'analyse sont imprédictibles, contrairement aux données qui sont définies lors de la conception de l'application. 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[Bim07].

Un exemple de requête multidimensionnelle portant sur l'application de Figure 1.6a 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.

Un panorama des opérateurs OLAP proposés dans la littérature. Les plus communs sont :

- **Pivoter (pivot, swap)** : Permet d'interchanger deux dimensions
- **Forer (drill-down)** : Permet de descendre dans la hiérarchie de la dimension. Ex.

BD Classique	OLAP
S-projection	Slice
S-sélection	Dice
S-agrégation	Roll-up
S-désagrégation	Drill-down
S-union	–

TABLE 1.2 – Correspondance entre les opérateurs de BDS et OLAP

visualiser le nombre d'accidents par mois au lieu de par année.

- **Remonter (drill-up, roll-up)** : Permet de remonter dans la hiérarchie de la dimension. Ex. visualiser le nombre d'accidents par année au lieu de par mois.
- **Forer latéralement (drill-across)** :
 - Permet de passer d'une mesure à l'autre. Ex. visualiser le coût des travaux au lieu du nombre d'accidents
 - Permet de passer d'un membre de dimension à un autre. Ex. visualiser les données de Laghouat au lieu de celles du sidi-makhlouf

Les opérateurs OLAP ont une similitude avec les opérateurs des bases de données statistiques : les opérateurs de l'algèbre relationnelle : «select», «project», «union» et «aggregate». Le tableau ci-dessous, résume cette correspondance :

1.3.5 Architecture des entrepôts de données

Les architectures des systèmes d'entrepôts de données sont classiquement des architectures à trois niveaux, comme montré dans la Figure 1.7, constituées par un entrepôt de données, un serveur OLAP et un client OLAP.

Une architecture d'entrepôt de données possède les caractéristiques suivantes :

- les données sources sont extraites de systèmes, de bases de données et de fichiers
- les données sources sont nettoyées, transformées et intégrées avant d'être stockées dans l'entrepôt
- l'entrepôt est en lecture seulement et est défini spécifiquement pour la prise de décision organisationnelle
- les usagers accèdent à l'entrepôt à partir d'interfaces et d'applications (clients)

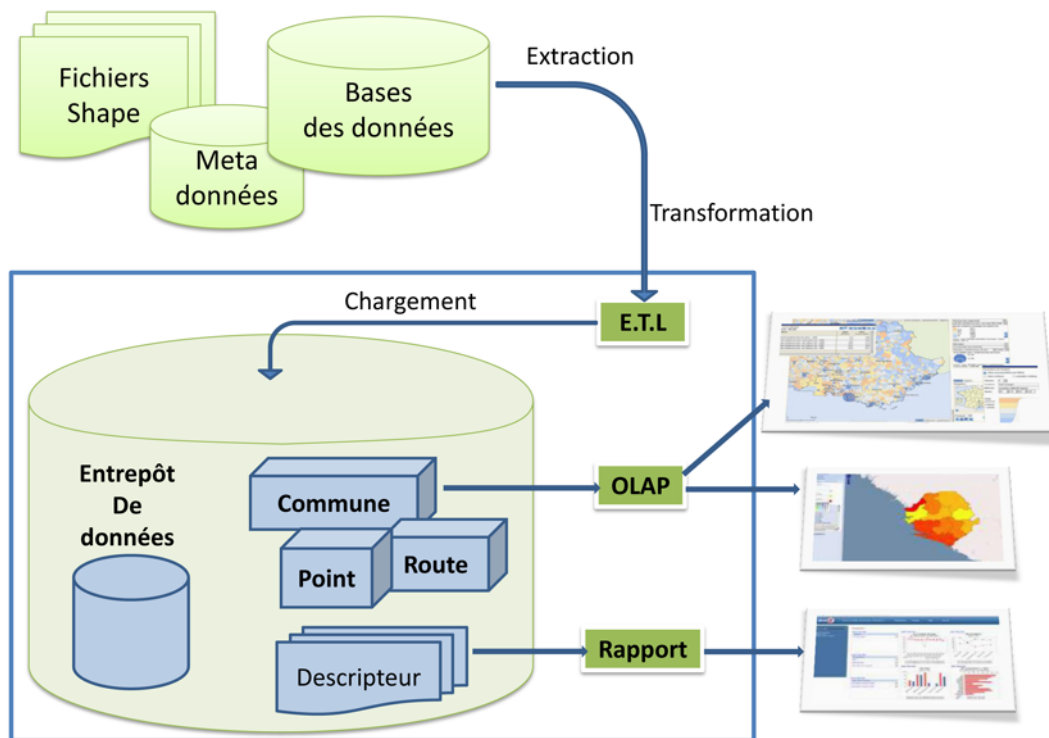


FIGURE 1.7 – Architecture à trois niveau d'un système d'entrepôt de données

1.3.5.1 Entrepôt de données

Le premier niveau est un SGBD (Système de Gestion des Bases de Données). Les données sont extraites à partir de différentes sources de données généralement sont des bases de données transactionnelles, ces données seront 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.[Bim07]

1.3.5.2 Serveur OLAP

Le deuxième niveau de l'architecture est un serveur OLAP, par exemple Mondrian, 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.

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.) voir 1.3.4. 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 [VAJ96].

Pour optimiser l'accès aux données agrégées, le serveur OLAP pré-calculé 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[Bim07].

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[Bim07].

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[Bim07].

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 per-

mettent par exemple de naviguer dans les hiérarchies et de définir des mesures dérivées complexes[Bim07].

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.

Implémentation physique d'un serveur OLAP On trouve principalement trois approches pour l'implémentation de serveurs OLAP, Multidimensional OLAP (MOLAP), Relational OLAP (ROLAP) et Hybrid OLAP (HOLAP). Les serveurs MOLAP extraient les données de l'entrepôt de données Spatiale 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 [Ral96] 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 [WIN98].

La modélisation logique d'une base de données multidimensionnelle selon l'approche ROLAP ne systématise pas l'utilisation de la 3^{ème} forme normale, contrairement aux systèmes OLTP. Le modèle logique le plus utilisé est le schéma en étoile [Ral96]. Le schéma en étoile (Figure 1.8a) 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 1.8b). 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 1.8c). Cette représentation logique permet d'utiliser l'opération de drill-across[Bim07].

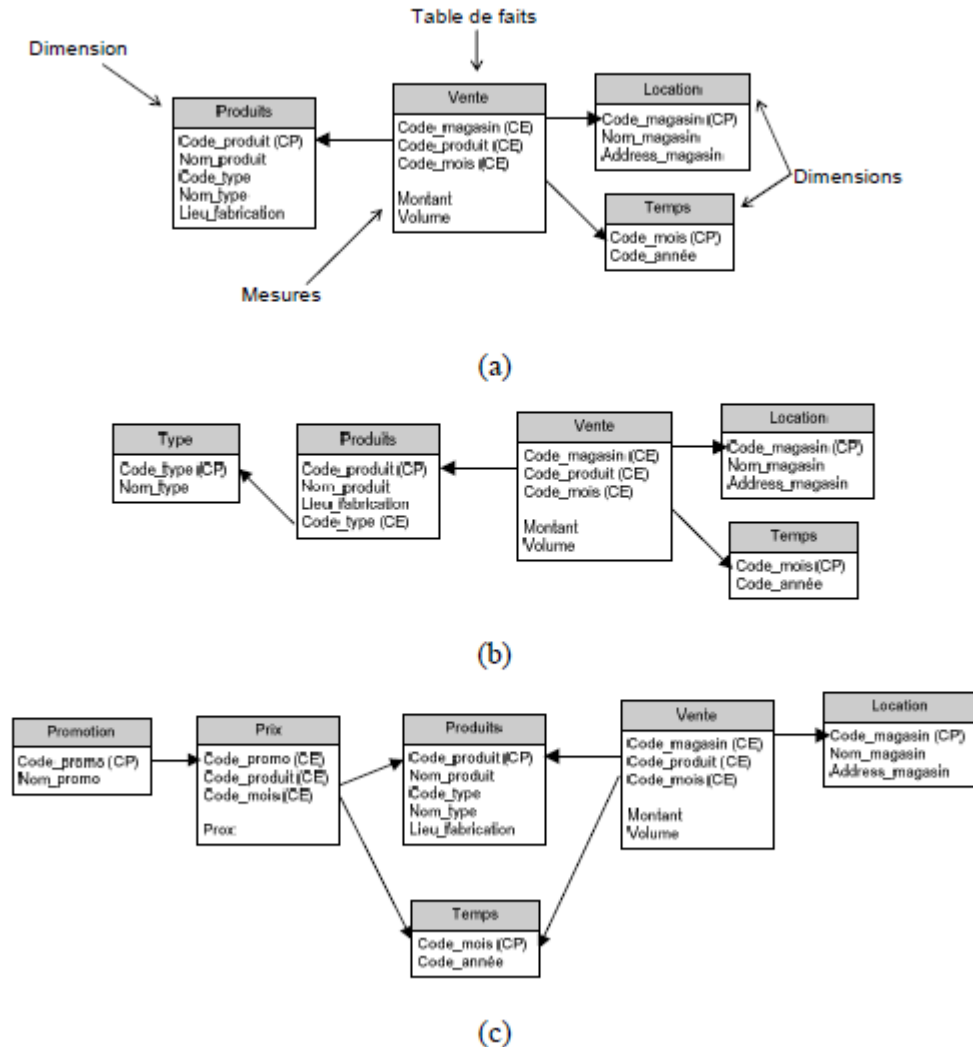


FIGURE 1.8 – a)Schéma en étoile b)Schéma en flocon c)Schéma en constellation [Bim07]

Il existe aussi 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.

1.3.5.3 Client OLAP

Le dernier niveau est un client OLAP, 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[Bim07].

1.4 Conclusion

Dans ce chapitre, nous avons présenté un panorama des concepts principaux des SIG et des EDS. Nous avons commencer par la description des caractéristiques principales de l'information géographique, ensuite nous avons donnés les fonctionnalités des SIG et nous avons mis la lumière sur la puissance acquise par ses systèmes lorsqu'ils sont fusionnés avec les technologies Web.

Également nous avons décrit les concepts à la base de l'analyse multidimensionnelle, structure de données et opérateurs, et nous avons détaillés l'architecture classique des systèmes d'ED et aussi les caractéristiques fondamentales des modèles formels pour les bases de données multidimensionnelles. C'est vers cette direction que se positionnent nos travaux concernant l'introduction de l'information spatiale dans les ED, comme décrit dans les chapitres suivants.

Chapitre 2

Travaux Connexes

Rappelons que nous étudions ces travaux dans un but d'apprentissage plus qu'une étude critique. Notre but est d'acquérir un savoir faire en terme de conception des entrepôts de données géographique.

Dans ce chapitre nous présentons différents travaux dans la littérature liées aux SIG et aux entrepôts de données, ainsi qu'à la prise en charge des différents concepts et caractéristiques de l'information géographique, avec les quels on va comparer nos travaux dans les derniers chapitres.

Dans le cadre de notre étude dans la littérature, nous avons choisi et sélectionnées 5 travaux parmi les plus récents :

1. Les travaux de Brisebois
2. Les travaux de Pestana et al.
3. Les travaux de Bimonte et al.
4. Les travaux de Bimonte
5. Les travaux de Selmoune et al.

Mais avant de commencer à citer ou à taxonomier ces travaux nous pouvons pas omettre celles du pionnier *Yvan Bédard*[BP89]. C'est le but de la section suivante.

2.1 Les Travaux de Bédard

La modélisation des bases de données géospatiales pour les applications SIG à toujours posé des défis pour les analystes et les développeurs des systèmes. Pour aider à résoudre ce problème, plusieurs logiciels et techniques de modélisation des BDD ont été proposées dans les 25 dernières années.

Cependant, le nombre de travaux liés aux ED et information géographique est timide. Et souvent ces travaux sont le produit d'une application spécifique et très peu de travaux ont ciblé le côté modélisation et méta-modélisation de méthodes formelles puissantes et génériques.

Dès l'an 1989, Bédard et al. se sont intéressés aux questions liées à la modélisation et l'extension de modèles des BDD aux informations spatiales ou géographiques. En effet, pour la première fois Bédard et Paquette[BP89] ont proposés une méthode formelle pour étendre l'UML avec des stéréotypes identifier par une notation textuelle ou de pictogrammes. Cette première solution a été testée dans plusieurs projets et amélioré avec le temps pour devenir Modul-R [MMB⁺96] [BPC92].

Un premier test a été fait en 1996-97 pendant la conception de modèle Orienté Objet du réseau routier du gouvernement de la Colombie Britannique. Dans l'été 1997, la première version de l'outil **Perceptory** et mise en marche ; il implémente PVLs (Plug-ins for Visual Languages) en tant-que stéréotypes UML.

En parallèle, autres chercheurs ont construit des variations de pictogrammes, ajoutant d'autres nouveaux pictogrammes ou même développer des concepts similaires. La Figure 2.1 de Filho et Lochpe présente une étude chronologiques de développement des formalismes. Le texte original de cette figure était en Portugais, nous avons fait la traduction vers le Français.

Comme légende de la Figure 2.1, on va donner la signification des trois formalismes car les abréviations entre parenthèses représentent les méthodes pour chaque chercheurs. Plus de détaille dans l'article de Filho et Lochpe [FI99].

ER Entité Relation

OO Orienté Objet

IFO Modèle sémantique formelle pour les base de données [AH87]

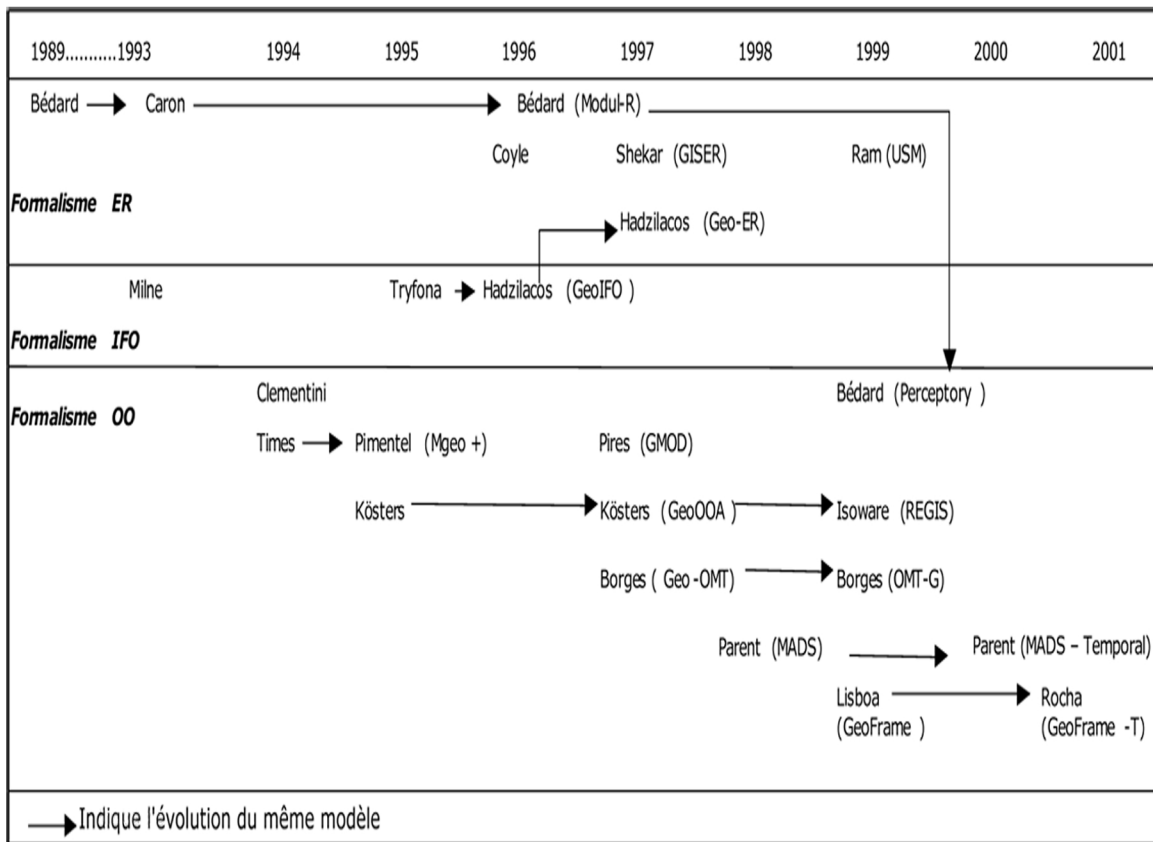


FIGURE 2.1 – Évolution chronologique des modèles conceptuels de données pour les SIG [FI99]

2.2 Les travaux de Brisebois 2002

En 2002, Alex Brisebois [Bri02] a mis le point sur les notions d'entrepôt de données et d'intégration de données géospatiales. il a tenté d'identifier, catégoriser et même de solutionner les problèmes types, d'ordre théorique et pratique liés à la mise en place d'un entrepôt de données spatial. L'auteur a rencontré un grand nombre de difficultés durant la réalisation de son prototype en utilisant comme benchmark un petit entrepôt de données sur la forêt de Monmorency. Il a classé ces difficultés comme suit :

Intégration sémantique :

- Validation de contraintes d'intégrité des données,
- Absence partielle de méta-données,
- Difficulté à suivre une évolution descriptive.

Intégration spatiale :

- Observation discrète de phénomène continu,
- Recherche d'entités persistantes c-à-d la possibilité de comparer les valeurs descriptives d'une entité territoriale sur plusieurs époques,
- Manipulation matricielle/vectorielle,
- Erreurs topologiques,
- Traitement géométrique,
- Lourdeur des jeux de données.

Entrepôt de données :

- Modélisation adéquate de l'entrepôt en fonction des besoins pour l'analyse spatio-temporelle,
- Comparaison des données détaillées (données brutes provenant de sources différents).

Parmi ces problèmes, nous nous intéressons par ceux qui portent sur l'intégration des données, Brisebois et al.[MBB02] ont traité des problèmes liés à la constitution de structures multidimensionnelles dont l'alimentation est faite à partir de données géospatiales hétérogènes existantes et présentant des problèmes d'intégration à la fois sémantique, temporelle et spatiale.

2.3 Les travaux Pestana & al. 2005

Pestana et al. proposent un modèle de données spatial multidimensionnel qui explore les fonctionnalités spatiales et temporelles fournies par un entrepôt de données spatial [PdSB05]. Leur approche de modélisation des données intègre uniformément les données

spatiales et non-spatiales, aussi elle se différencie de la modélisation de données traditionnelle.

Par le mot **Intègre**, ils entendent non seulement stocker des données spatiales et non-spatiales dans la même base de données, mais également de stocker constamment les topologies concernant chaque changement géométrique au fil du temps.

De plus, dans leur approche de modélisation de données, ils étendent le schéma en étoile traditionnel au domaine spatial en insérant des types standards de données spatiales et des capacités de traitement spatial dans l'entrepôt de données spatial.

De plus, ils développent un outil de modélisation visuel dédié à la modélisation des données spatiales nommé *Perceptory*[BLjPN04]. Ils exploitent également les fonctions d'agrégations natives qui existent dans les bases de données spatiales pour récapituler les objets géographiques et pour réduire la complexité des requêtes sous-jacentes.

Cependant, les chercheurs ont omis de traiter des aspects tels que les rôles d'objets, contraintes, la qualité des données, les fonctionnalités de SOLAP et ETL spatial telsque l'extraction, transformation et le chargement, pour un entrepôt de données de multi-granularité. En effet aspects doivent être pris en considération pour les propriétés spatiales, afin d'éviter l'incohérence et l'ambiguïté.

2.4 Les travaux de Bimonte & al. 2006, 2007

Plus récemment, Bimonte et al. ont proposé une double contributions. D'une part, ils ont défini le concept de *Dimension géographique* comme une dimension dont les composantes sont des objets géographiques du monde réel (une ville, un bâtiment) et pour lesquels une représentation cartographique peut exister. D'autre part, ils ont spécifié trois types d'hierarchies possibles qui permettent de prendre en compte la sémantique de différents relations qui peuvent exister entre les composantes géographiques de différents niveaux [BTMR07].

Ensuite, leurs différentes contribution a abouti à la naissance du prototype **GeWolap** [BWTM06], un outil Web, qui intègre des fonctionnalités OLAP et SIG (SIG-OLAP) dans un environnement unique qui synchronise les différentes composantes cartographiques et tabulaires afin de compenser le manque ou bien les limitations dans le support des mesures géographiques ou bien les problèmes liées a l'introduction des mesures géographiques dans l'OLAP spatial tel que défini dans [BTM06].

L'outil GeWolap permet une navigation interactive dans un hypercube conçu suivant le modèle "GeoCube" [BTM06]. L'architecture du prototype, tel que décrit par les auteurs

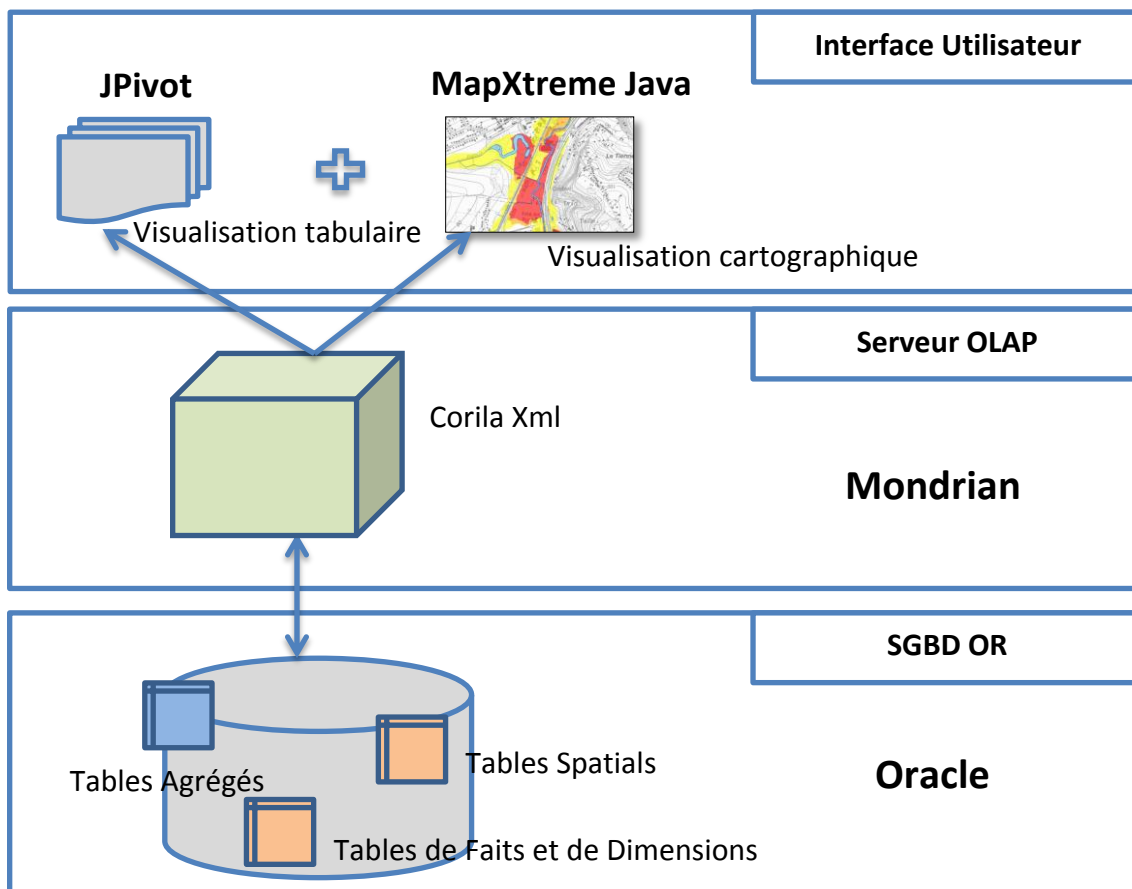


FIGURE 2.2 – Architecture de GeWOlap
[Bim07]

est composé de trois niveaux :

1. un SGBD objet relationnel supportant les données spatiales,
2. un serveur OLAP pour la gestion des requêtes multidimensionnelles,
3. une interface client (front-end) basée sur un client Web-OLAP et un client Web-GIS.

La Figure 2.2 illustre cette architecture.

Les Auteurs de ce travail se sont concentrés sur l'intégration des mesures spatiales ou géographiques et l'introduction d'autres fonctionnalités SIG comme par exemple l'analyse de la mémoire tampon (buffer), le calcul de la surface ou des outils de reclassification.

Nous devons insister sur le fait que l'intégration des mesures géographique dans l'OLAP spatial en d'autre terme la combinaison SIG-OLAP soulève plusieurs problèmes de point de vue théorique et implémentation. Par exemple une double difficultés qui se présente

est : La prise en compte d'une gestion dynamique de la structure de l'hypercube, en termes de mesures et de hiérarchie, et le calcul des mesures associées aux nouveaux membres.

2.5 Travaux de Bimonte 2008

Dans un autre travail [Bim08], Bimonte s'est intéressé à l'introduction de la composante sémantique de l'information géographique et la flexibilité de l'analyse spatiale dans les systèmes Spatial OLAP. L'auteur propose, en plus du modèle formel **GeoCube**, une algèbre associée qui prend en compte les concepts principaux de l'OLAP géographique, ce dernier est un nouveau paradigme d'analyse multidimensionnelle introduit par le chercheur. Bimonte redéfinit le concept de la mesure (notion vu dans la section 2.3.2) comme un objet géographique caractérisé par plusieurs attributs thématiques et par un attribut spatial.

Ce modèle présente deux limites importantes mentionnées par l'auteur lui même. En effet, GeoCube ne prend pas en compte d'une part les relations topologiques des hiérarchies spatiales dans le processus d'agrégation. D'autre part, il ne gère pas le type d'attribut (additif, semi-additif et non-additif) dans la définition des fonctions d'agrégation.

2.6 Travaux de Selmoune & al. 2012

Dans le cadre d'une application spécifique [SAA12], les auteurs ont proposé une approche de conception du modèle spatial multidimensionnel qui est une adaptation de la méthodologie proposée dans [KRRW00].

Il s'agit d'une approche progressive basée sur cinq principales étapes (Figure 2.3), ils ont décrit ces étapes en les appliquant à la gestion des ressources en eau pour le compte de l'Agence Algérienne des Ressources Hydraulique (ANRH).

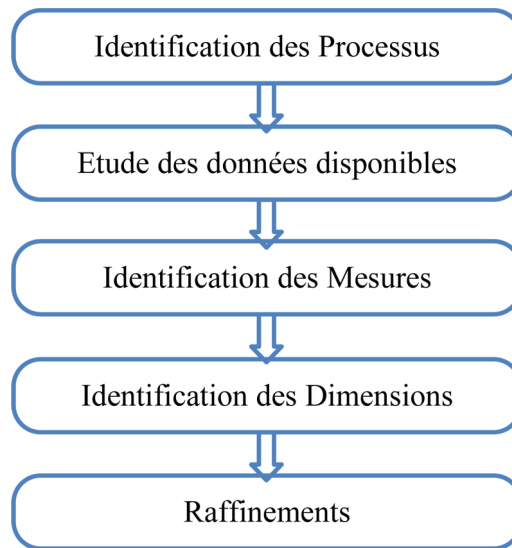


FIGURE 2.3 – Processus de conception multidimensionnelle

Le résultat de leur conception est un modèle multidimensionnel comme illustré dans Figure 2.4 qui met en évidence les mesures, les dimensions spatiales et non-spatiales qui sont pertinentes dans un processus de décision.

Ils ont utilisé également dans leur outil les opérateurs OLAP pour analyser les mesures à travers différents axes.

Quoique, ils délivrent un entrepôt de données géographique, mais ce travail est très peu documenté, et très peu de détail sur sa conception et aussi sur comment ils ont géré les enjeux liés à l'intégration des données spatiales dans un tel environnement.

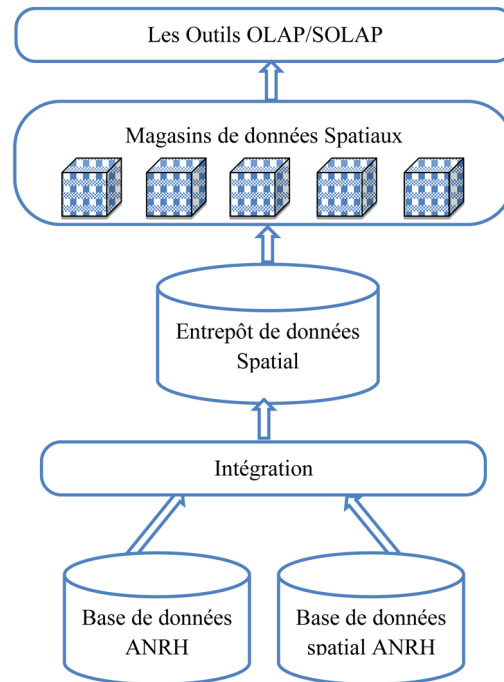


FIGURE 2.4 – L’architecture de la solution ANRH

2.7 synthèse, comparaison et discussion

	Méthodologie de conception	Outil développé	Web / Desktop	Composante de l’info géographique intégré
Bédard et al. 1989-2002	Objet relationnel En étoile En flocon de neige	Perceptory v1.0	Desktop	Spatiale
Brisebois 2002	En constellation	–	–	Spatiale Sémantique
Pestana et al. 2005	TAD spatial (PVL)	Perceptory v2.0	Desktop	Spatiale Topologique
Bimonte et al. 2007	En flocon de neige	GeWolap	Web	Spatiale
Bimonte 2008	Oracle spatial En flocon de neige	–	–	Sémantique
Selmoune et al. 2012	Multidimensionnel Traditionnel	Solution ANRH	Desktop	–

Nous avons étudié un ensemble de travaux parmi les plus caractérisait et récents qui traitent des problèmes liées à l'intégration d'une ou plusieurs composantes de l'information géographique aux BDD et ED.

Dans la Table 2.1, nous résumons et synthétisons ces travaux on prenant comme critères de présentation et comparaison la méthodologie suivie lors de la conception, quelle composante d'information spatiale est touchée, développement d'un outil ou pas ? et sur quelle environnement leurs outil fonctionne-t-il ?.

Cette étude nous a permis d'avoir une idée sur le domaine des Entrepôts de données géographiques, et d'en tirer certaines conclusions :

- Le domaine des EDG est un domaine très récent avec très peu de travaux.
- Les méthodes et les outils de conception des systèmes à base d'EDG restent incomplets et limités.

Suite à cette étude qui nous a permis de rassembler un ensemble d'éléments nécessaire pour entamer le monde du géodécisionnel, nous allons proposer notre approche de modélisation et conception du système projeté à savoir la conception d'EDG et la solution Web-SIG qui sera accessible aux utilisateurs via le Web. C'est l'objet de chapitre suivant.

Chapitre 3

Conception

Dans le chapitre précédent nous avons présenté un panorama de travaux traitant l'intégration de l'information géographique dans les entrepôts de données. La méthodologie suivie se diffère d'un travail à l'autre à savoir la modélisation de données en étoile, en flocon de neige ou en constellation de faits.

Ces travaux se sont souvent focalisés sur l'intégration de la composante spatiale de l'information géographique seulement, et n'offrent qu'une vue générale ou superficielle de leurs systèmes.

Parmi les principales objectifs du projet PNR dans lequel ce travail est inscrit, est de proposer une approche de conception des entrepôts de données géographiques basées Web.

Nous proposons dans ce chapitre la démarche suivie pour modéliser les données géographiques et tabulaires qui vont alimenter l'entrepôt du futur système appelé "GEESlagh" (Gestion Socio-Spatial pour Laghouat). Et mener à bien la conception de l'entrepôt de données géographique.

Nous commençons par une étape de modélisation des données multidimensionnelles, puis la validation du modèle élaboré, et en finissons par le chargement ou l'intégration des données dans l'EDG.

Vue la difficulté d'avoir des données complète sur toutes les domaines pour enrichir le système, le contact avec des experts dans le domaine du bâti nous a permis d'établir un processus métier pour ce domaine d'étude, et d'avoir un ensemble de données cartographiques suffisantes pour tester notre système.

3.1 Contexte : La gestion socio-spatiale en milieu semi-aride ; cas de la wilaya de Laghouat

Ce travail rentre dans le cadre d'un projet national de recherche (PNR). Il s'agit de produire un outil d'aide à la décision pour tous les acteurs socio-économique de la wilaya de Laghouat, intégrant et synthétisant les données socio-spatiales (images satellitales de 1972 à 2007 complétées par des levés in situ) et données ONS (Office National des Statistiques), à l'échelle de la commune, complétées par des enquêtes sur les chefs lieu de communes.

A fin de réaliser la totalité de la solution cible de ce projet, le travail a été partagé selon les différents savoir-faires des membres de l'équipe PNR tout en maintenant et préservant une étroite collaboration :

1. Les informaticiens se sont occupés de la conception et l'implémentation de la solution au sein d'un environnement conçu aux normes du génie logiciel.
2. Les spécialistes en télédétection et aménagement des territoires ont réalisé la partie extractions et synthèses des données socio-spatiales à partir d'images satellitales.
3. Les sociologues et géographes se sont chargés de la collecte de données socio-économiques réelles extraites d'études de terrains, des enquêtes et des études statistiques.

Ce qui nous a été confié est plus spécialement :

- Rassembler les savoirs faire de l'informatique décisionnelle, des entrepôts de données géographiques et du Web-mapping.
- Atteindre le principal objectif de notre travail, c'est de proposer une méthode générique de conception des entrepôts de données géographique basée sur des visualisations sur le Web.
- Appliquer cette méthode à notre cas : Évolution du *Bâti*)

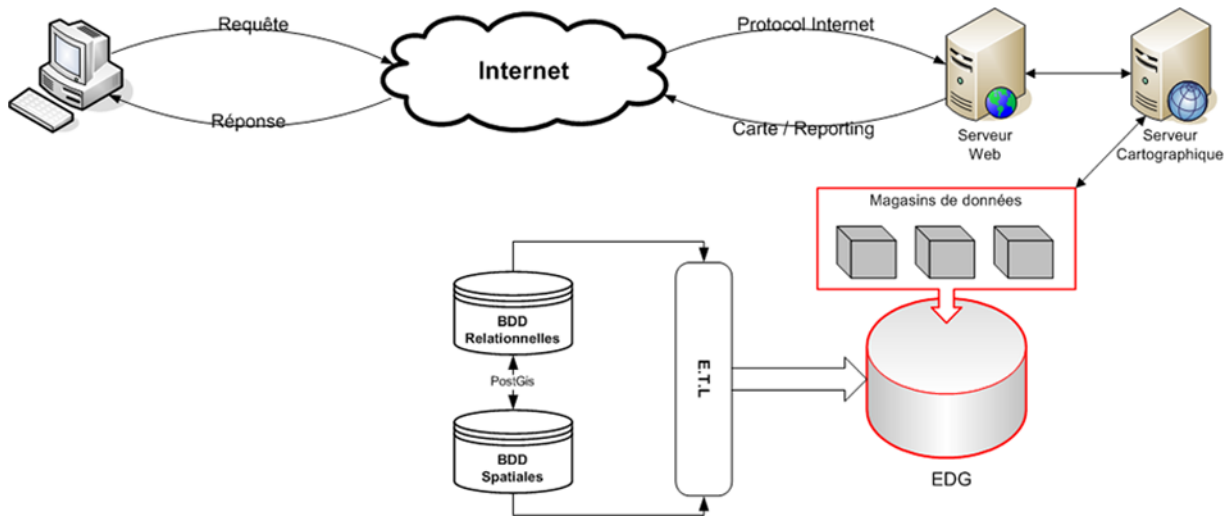


FIGURE 3.1 – Vue générale du système projeté

3.2 Entrepôt de données géographique dans le Web

La figure 3.1 schématise d'une façon sommaire le système Web-EDG envisagé. Le but est de répondre selon le domaine choisi à une requête complexe à la fois spatiale et temporelle ; on donne à l'utilisateur la possibilité de faire une comparaison entre plusieurs époques de la même zone cartographique.

Le processus commence par un simple click (navigation sur le site), derrière ce click sera déclenchée une requête spatiale complexe transmise via Internet en passant par un serveur Web vers un serveur cartographique. Le rôle de ce dernier est de décoder cette requête puis interroger l'entrepôt (en couleur rouge). Pour notre travail, nous nous sommes intéressés à la conception et réalisation d'EDG qui permet de répondre à tous types de requête spatiales y compris les requêtes tabulaires.

3.2.1 Modélisation dimensionnelle des données

Les modèles proposés en littérature (Pestana et al. 2005) (Bimonte 2008) (Selmoune et al. 2012) qui introduisent l'information géographique comme axe d'analyse, sont caractérisés tous par l'orienté métier, c-à-d l'outil résultant de leurs modélisation est destiné à un domaine bien précis.

L'élaboration d'une méthode générique pour la conception d'un système de cartographie en ligne alimenté par un entrepôt de données géographique, est une phase primordiale

dans notre travail.

En plus des standards de conception des entrepôts de données proposés par Inmon [W.H96] et Kimball [Ral96], notre méthode introduit des nouvelles étapes dans le processus de conception dans le but de rendre le système à la fois extensible et fiable en terme de satisfaction des requêtes des utilisateurs dans tous les domaines envisagés.

On résume ces étapes comme suit :

- Le choix du processus de modélisation,
- Le regroupement des modèles dimensionnelles en étoile en un seule modèle en constellation de faits,
- Le choix des dimensions,
- L'identification des faits et l'intégration des données.

Pour mieux concevoir et tester le système nous avons établie un ensemble de requêtes utilisateurs pour le processus métier pour le domaine du bâti et son évolution.

Requête 1 *Quelle est la surface bâtie de la zone X pour l'année A et l'année B ?*

Requête 2 *Quelle était l'état du bâti dans la zone X pour l'année A ?*

Requête 3 *La comparaison du bâti entre deux époques pour la même zone géographique*

3.2.1.1 Le choix du processus de modélisation dimensionnelle

Le processus de production des résultats est constitué d'ensemble des informations portant sur plusieurs domaines. Toutes les données ne sont donc pas traitées de la même manière : sur chacun de ces domaines, les variables mesurées et estimées diffèrent.

Les variables mesurées et leurs sources dépendent fortement du domaine d'étude choisi. Le tableau 3.1 montre quelques cas envisagés. Par exemple, seules des données de statistiques issue de la Monographie (rapport annuel de l'ONS) seront disponibles pour le domaine d'étude socio-économique. A l'inverse de nombreuses variables seront disponibles pour le Transport.

Due aux natures différentes d'opérateurs, nous sommes contraints de concevoir pour chaque domaine d'étude un modèle différent. La figure 3.2 résume les modèles dimensionnels choisis (chaque domaine est modéliser sous forme d'étoile, comme représente l'info-bulle flottante du domaine "commune" dans le premier niveau). ces modèles en étoile vont former l'ensemble du modèle dimensionnel de l'EDG.

Domaine d'étude	Variables
Commune	Surface
Transport (réseau routier)	Routes, Chemins, Taux de trafic,...
Sol	Surface, volume, taux de changement,...
Socio-économique	Population, croissance,..
Tourisme	Sites historiques, patrimoine culturel,...
Démographique	Population,...
Éducation	éducation-var
Météorologique	météo-var
Sport et loisirs	sport-var
Santé	santé-var

TABLE 3.1 – Les domaines d'étude

Par définition, un entrepôt de données en général et l'EDG en particulier, est orienté sujet. De ce fait notre entrepôt va regrouper plusieurs sujets ou métiers ; chaque domaine d'étude représente un métier. C'est ce qu'on appelle dans l'informatique décisionnel "*les magasin de données ou d'informations*" (Data mart).

Ce schéma lie en hiérarchie les différents modèles de données dimensionnels constituant l'ensemble du système d'entrepôt de données géographique.

Ainsi, le modèle dimensionnel «Commune» rassemble toutes les informations de type spatial (zones, agglomérations, cités, ...) puisque ces informations sont collectées sur l'ensemble des divisions territoriales. Au deuxième niveau, les modèles «étude de sol», «réseau routier», «tourisme» et «socio-économique» contiennent toutes les informations levées au niveau de la commune correspondante. Finalement, les modèles du troisième niveau couvrent l'ensemble des informations relatives à un domaine d'étude du deuxième niveau.

La diversité des domaines d'étude ainsi que leurs relations dans l'hiérarchie, nous amène à des volumes de données de tailles différentes ; dans le modèle «Commune», toutes les lignes permettant de calculer des surfaces à partir des points cartographiques vont être stockées, qu'ils aient été visualisés ou pas. En revanche, les modèles du deuxième niveau, comportent uniquement des données correspondantes aux régions visualisées et ils sont de plus associés à d'autres données tabulaires.

Une chose importante dans cette première phase de modélisation c'est d'avoir remar-

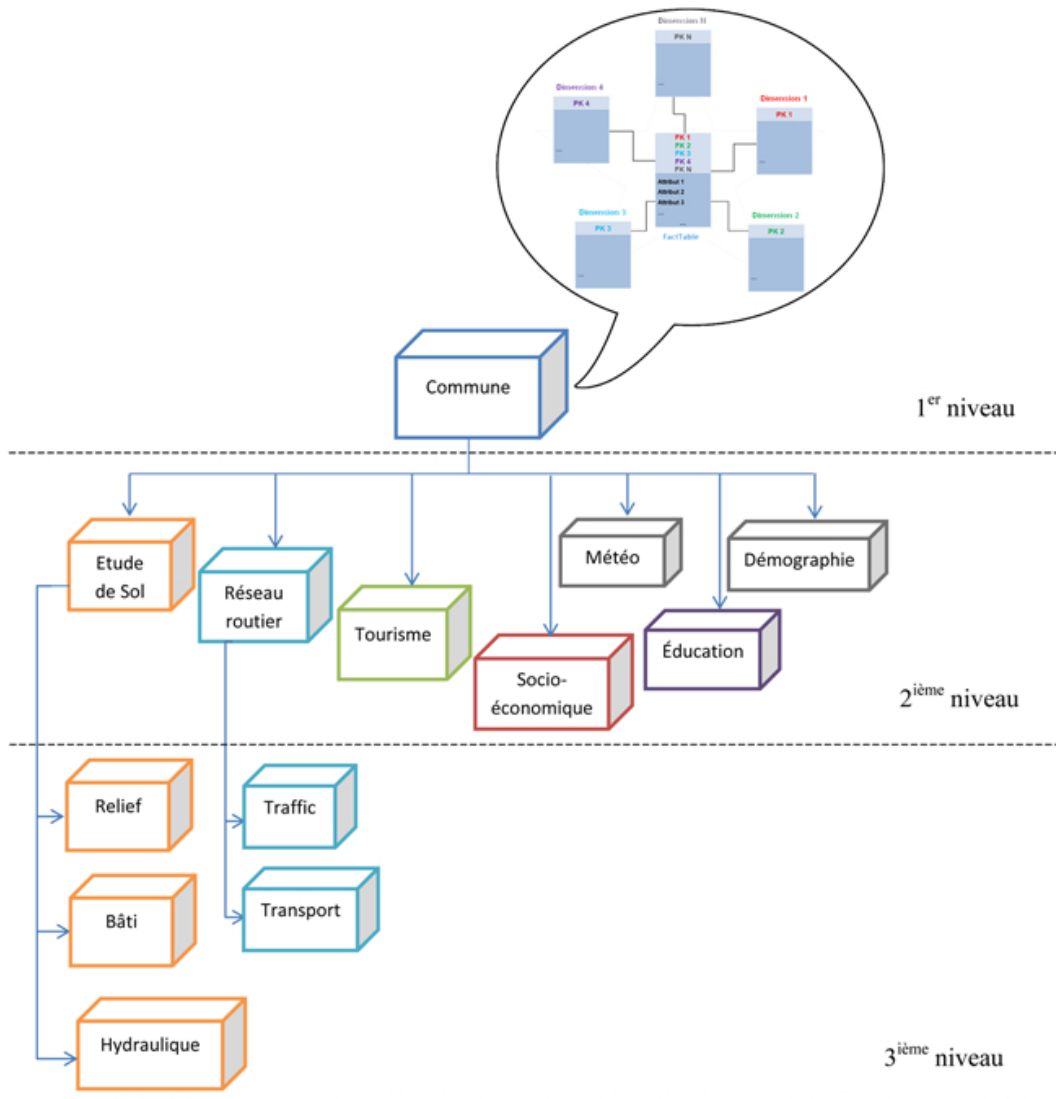


FIGURE 3.2 – Modèles dimensionnels

quer que de nombreuses dimensions allaient se partager. Par exemple dans les dimensions cartographiques presque toutes les dimensions vont se joindre à chaque modèle dimensionnel envisagé. De ce fait, une modélisation en constellation de faits est optimale pour notre modélisation.

3.2.1.2 Constellation de Faits

Lors de la modélisation des dimensions communes, il faut s'assurer qu'elles ont toutes la même similarité, elles utilisent les mêmes clés, les mêmes noms de colonnes d'attributs et les mêmes définitions et valeurs d'attributs, ce qui garantira des regroupements d'informations cohérents.

À la fin de cette première phase de modélisation dimensionnelle des données, plusieurs schémas en étoile ont été envisagés et regroupés dans un seul modèle en constellation représentés dans la figure 3.3.

Le cadran en rouge montre le fragment du modèle correspond au domaine du bâti. Ainsi le dictionnaire de données associé à ce modèle est présenté dans l'annexe A.

Pour clarifier la jointure entre les différentes étoiles constituant le modèle en constellation de faits, on propose la schématisation illustrée dans la Figure 3.4 :

En effet, les dimensions cartographiques, statistiques et temporelles, sont des dimensions communes entre les différents modèles en étoile envisagés, nous nous intéressons qu'aux dimensions spatiales et temporelles pour notre cas (le bâti).

3.2.2 Validation du Modèle

À ce stade de modélisation, un modèle en constellation de faits est élaboré et doit être soumis à un ensemble de critères afin de prouver sa validité.

Dans la littérature on trouve pas une méthode ou un standard spécialement dédié à la validation des modèles dimensionnelles. Un modèle dimensionnel est dit valide s'il répond aux besoins soulignés lors de l'élaboration du processus métier. Nous allons prouver la validité de notre approche de modélisation tout en suivant les étapes suivantes :

3.2.2.1 Niveau de granularité

La déclaration de niveau de granularité est une étape critique après la modélisation du processus, car si un grain trop élevé est choisi, le modèle se limite à un nombre plus faible

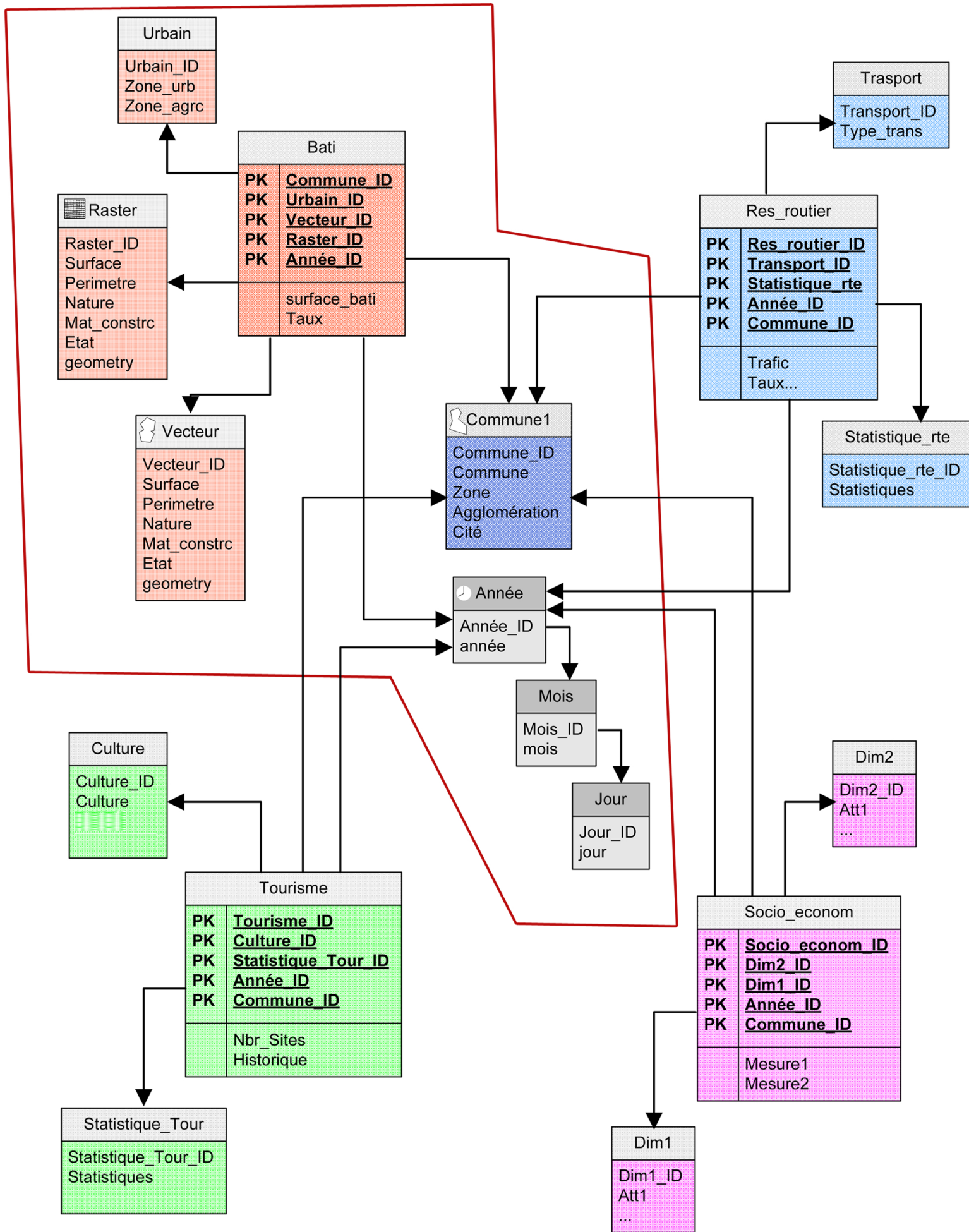


Figure 3.3 - Vue partielle du modèle en constellation de faits

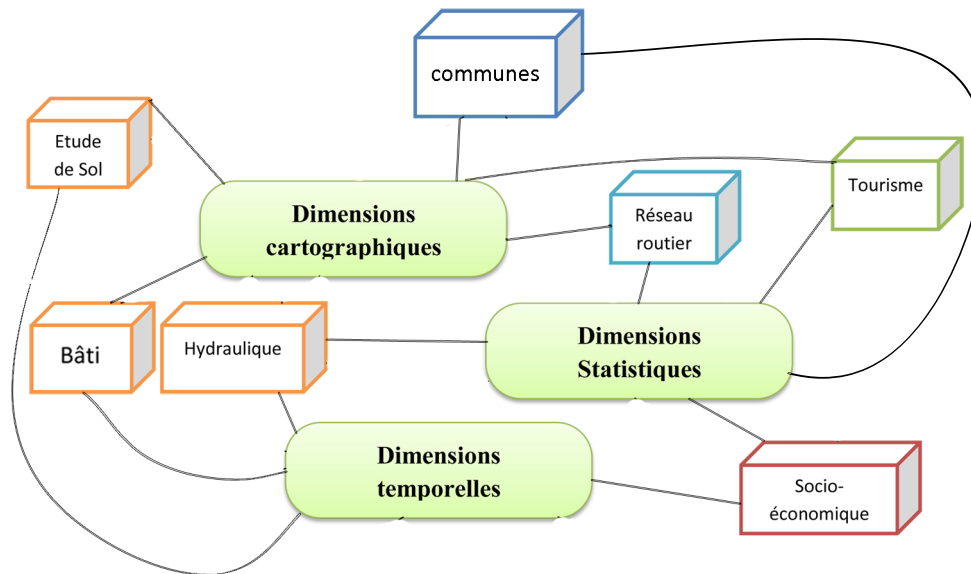


FIGURE 3.4 – Vue simplifiée du Modèle en constellation de faits précédent

de dimensions ou à une information moins détaillée [W.H92]. Un modèle trop grossier est vulnérable rapidement à des demandes inattendues des utilisateurs qui voudraient faire un forage plus profond. Cependant enregistrer les données cumulées joue un rôle important dans l'optimisation des performances. Au contraire, ce qui est vivement recommandé c'est d'enregistrer les données les plus atomiques ; elles offrent une grande souplesse.

Plus clair, il faut spécifier exactement ce que représente un tuple de chacune des tables de faits. Par exemple le modèle «Bâti» dans lequel les mesures concernées sont la surface et le comptage des points. Il faut spécifier le niveau de détail de ces mesures, c'est-à-dire le plus petit niveau d'information que l'utilisateur voudrait. Nous allons focaliser sur la mesure surface.

En plus des dimensions statistiques et dimensions cartographiques, la dimension temporelle «Temps» est une dimension très importante pour pouvoir répondre à la requête temporelle. En effet, à cause de cet axe toutes les mesures ne sont pas additives, mais semi-additives.

A titre illustratif, afin d'estimer les superficies du bâti il faut effectuer au final une moyenne temporelle simple sur la période voulue. Par exemple si l'on cherche à estimer la surface en zone quelconque du bâti [1987-2010], il faut effectuer une moyenne des

estimations de ces surfaces sur chacune des années soit :

$$S_{[1987-2013]} = \frac{S_{1987} + S_{2001} + S_{2006} + S_{2013}}{4} \quad (3.1)$$

3.2.2.2 Hétérogénéité

Pour le moment les sources de données qui vont être utiliser pour alimenter l'EDG sont inconnues voir inexistantes, de cet effet il ne faut pas négliger les problèmes d'intégration des données provenant de différentes sources à différents moments. Au fil de temps de nouvelles données arrivent et de nouveaux protocoles mis en œuvre créant potentiellement une hétérogénéité sémantique, temporelle et spatiale. la structure des données peut être évolutive ce qui se traduit notamment par des définitions de variables qui changent, certains données sont négligeables d'autres sont nouvellement prises. le découpage des zones cartographique évoluent aussi et des nouvelles cartes sont générées.

L'organisation des données au sein des modèles dimensionnels est un peu particulière, les dimensions sont statiques et seules les tables de fait reflètent l'aspect dynamique. nos travaux porte sur la résolution de cette problématique d'intégration de ces informations évolutives dans ces dimensions et plus particulièrement ceux qui concerne la composante géographique.

L'intégration de nouvelle données ne produit pas des difficultés si la granularité a bien été définie. une opération de rechargement est nécessaire si la granularité est trop grande. Par exemple en ajoutant une nouvelle donnée issue d'une composition de données déjà existantes (le Taux d'accident sur le chemin x dans la période P sous la météo M), le modèle est ajusté pour tenir compte de cette données mais aucun chargement n'est nécessaire. par contre si une donnée est nouvellement mesurée alors il faut envisager un nouveau chargement.

Pour les données qui changent de définition, mais pas de plage de valeur comme les données discrètes il n'y a pas d'impact pour leurs intégration dans l'ED. C'est une information à communiquer aux utilisateurs. par exemple le Taux de trafic routier rentre dans cette catégorie. Le calcul du Taux de trafic repose sur quatre données la longueur et le type de la route (données statique) et nombre de véhicules et l'état de la route (données relevées sur le terrain). Alors au fil du temps les tranches de valeur de cette données restent "faible", "moyen", "élevé" et "très élevé". la dimension de l'entrepôt qui s'occupe de ce type de données reste donc identique.

Pour les valeurs des données évolutives dans le temps, une solution consisterait à faire évoluer les dimensions elles-mêmes en ramenant les données à une version unifié, cela apporterait une uniformité et une homogénéité des données de ces nouvelles dimensions,

en revanche il peut créer une perte de quantité d'informations importante.

En ce qui concerne l'hétérogénéité spatiale, l'usage qui en est fait dans l'ED peut comme il ne peut pas apporter de problèmes fonctionnels. Les informations spatiales de l'entrepôt qui évoluent concernent plusieurs propriétés. Prenant comme exemple la propriété du contour géographique des différents membres de la dimension cartographique. cette propriété n'a qu'un usage de représentation, si on choisi par exemple de stocker le contour géographique de la commune de « Sidi Makhlouf » de l'année 1984 dans la dimension « Communes » l'affichage des résultats des superficies de Bâti de 1987 du Sidi Makhlouf seront dessinées sur ce contour ce qui n'est qu'une illustration de l'information.

3.2.2.3 Choix des dimensions

Après l'identification des modèles, il reste à les enrichir avec un ensemble de dimensions représentant toutes les descriptions susceptibles de prendre des valeurs particulières pour chaque mesure. Pour les modèles définis précédemment plusieurs dimensions sont répertoriées. La dimension temporelle qui a un rôle très particulier, la dimension spatial qui comportent un attribut géométrique et les dimensions classiques textuelles qui regroupent les information statistiques de différents domaines.

Le manque de données et l'inexistence des informations géographiques pour des données attributaires dans certains domaines, rend notre travail un peut difficile. De ce fait, nous avons étudier seulement un fragment (Magasin de données) qui rentre dans le domaine d'étude du sol "le Bâti".

Dimension thématique

On va décrire l'ensemble des dimensions et mesures participantes dans le schéma multidimensionnel du bâti. Commençons par décrire les dimensions les plus classiques qui comportent les attributs textuels distincts, elles sont répertoriées en trois catégories :

- ◇ Les dimensions "Commune" qui vont caractérisées toutes les informations de territoire. Elles sont liées au zone du bâti, elles peuvent être spécifiques à un domaine d'étude ou bien communes à quelques domaines d'études.
- ◇ Les dimensions "Données bâti" comportent aussi des informations urbaines dite de territoire mais au niveau zone ou agglomération (zone bâti). certaines données sont partagées entre Données bâti et Commune.
- ◇ Les dimensions cartographiques qui vont caractériser les informations de reconnaissance de bâti. Elles sont issues du images de résolution et de dates d'acquisition

différentes (Des images Landsat : ETM+ de 1987, 2001 et 2006 de résolution 30 mètre) traitées avec des indices de bâti¹. les données de ces dimensions sont importées à partir d'un croisement réalisé en amont du chargement de l'entrepôt.

Dimension spatiale

La caractéristique principale de la dimensions spatiale c'est la présence d'un attribut géométrique dans les membres des différents niveaux. Dans la modélisation traditionnelle la dimension spatiale est traitée comme une dimension descriptive avec une représentation hiérarchique, dans notre modèle on attribues à chaque niveau dans les hiérarchies spatiales une définition géométrique (contour géographique). Chaque niveau représente une information géographique différente ; différentes granularités de l'information géographique. comme les dimensions classiques d'un modèle dimensionnel, les dimensions spatiales sont agrégatives hiérarchiquement

- ◇ «Dimension Communale» permet une analyse de faits suivant le découpage administrative. elle se décompose en quatre niveaux hiérarchiques :
 - Niveau "Commune" : contenant le polygone de la commune entière.
 - Niveau "Zone" : contenant les polygones des zones a l'intérieur de la commune.
 - Niveau "Agglomération" : contenant les polygones des agglomérations urbaines dans chaque zone.
 - Niveau "Cité" : contenant les polygones de chaque cité.
- ◇ «Dimension Urbain» dont la structuration en zone urbaine permet l'analyse dans le cadre de l'étude du bâti. Cette dimension se décompose en trois niveaux hiérarchiques :
 - Niveau "Commune" : contenant le polygone de la commune entière.
 - Niveau "Zone Urbaine" : contenant les polygones des zones urbaines.
 - Niveau "Zone agricole" : contenant les polygones des régions agricoles.

1. Mme Benkouider.F : Analyse diachronique du bâti

Dimension Temporelle

Chaque entrepôt de données a au moins une dimension temps, l'importance de cette dimension réside dans le suivi des changements de dimensions qui évoluent dans le temps. Tandis que les informations relative au bâti ont une lente changement au fil du temps, nous ne gardons que la colonne "Année".

Un objective majeur dans notre travail est basé sur la présence d'un axe de temps. Donc, ce que nous cherchions est de fournir un outil de cartographie sur le Web qui offre la possibilité de naviguer voir de revenir dans le temps. Nous sommes arrivés à établir cet objectif en automatisant un axe de temps " **Timeline** " qui associé chaque changement dans la dimension cartographique avec la période correspondante dans la dimension temporelle.

3.2.2.4 Identification des faits

Pour clôturer la modélisation multidimensionnelle de notre système, L'ultime étape c'est l'identification des mesures qui apparaîtront dans les tables de faits (dans le cas du bâti une seule table de faits). Pour le modèle du premier niveau, pour l'instant seule la surface est prise comme mesure. Ainsi pour le bâti on prend comme mesure l'évolution de la surface du bâti dans le temps comme indiqué dans la figure 3.5

3.2.3 Intégration des données

Après l'activité du modélisation, vient celle de l'ETC (Extraction, Transformation et Chargement) afin d'alimenter l'entrepôt de données. nous décrirons comment nous avons procédé afin d'extraire les données spécialement les information géographiques que ce soit raster ou vecteur en provenance des sources différentes la majorité sous forme de cartes. puis comment nous les avons structurées et sélectionnées au sein de notre entrepôt.

a) Préparation des tables de dimensions L'*extraction* c'est la première phase de l'ETL qui consiste à préparer les données brutes et de les adapter au modèle dimensionnel. Combiner les données, régler la qualité, identifier les mise a jour, construire les agrégats et traiter les erreurs. toutes ces actions forme la phase de *transformation*. La dernière phase celle de *chargement* est la plus simple a exécuter puisque les données ont été correctement préparées.

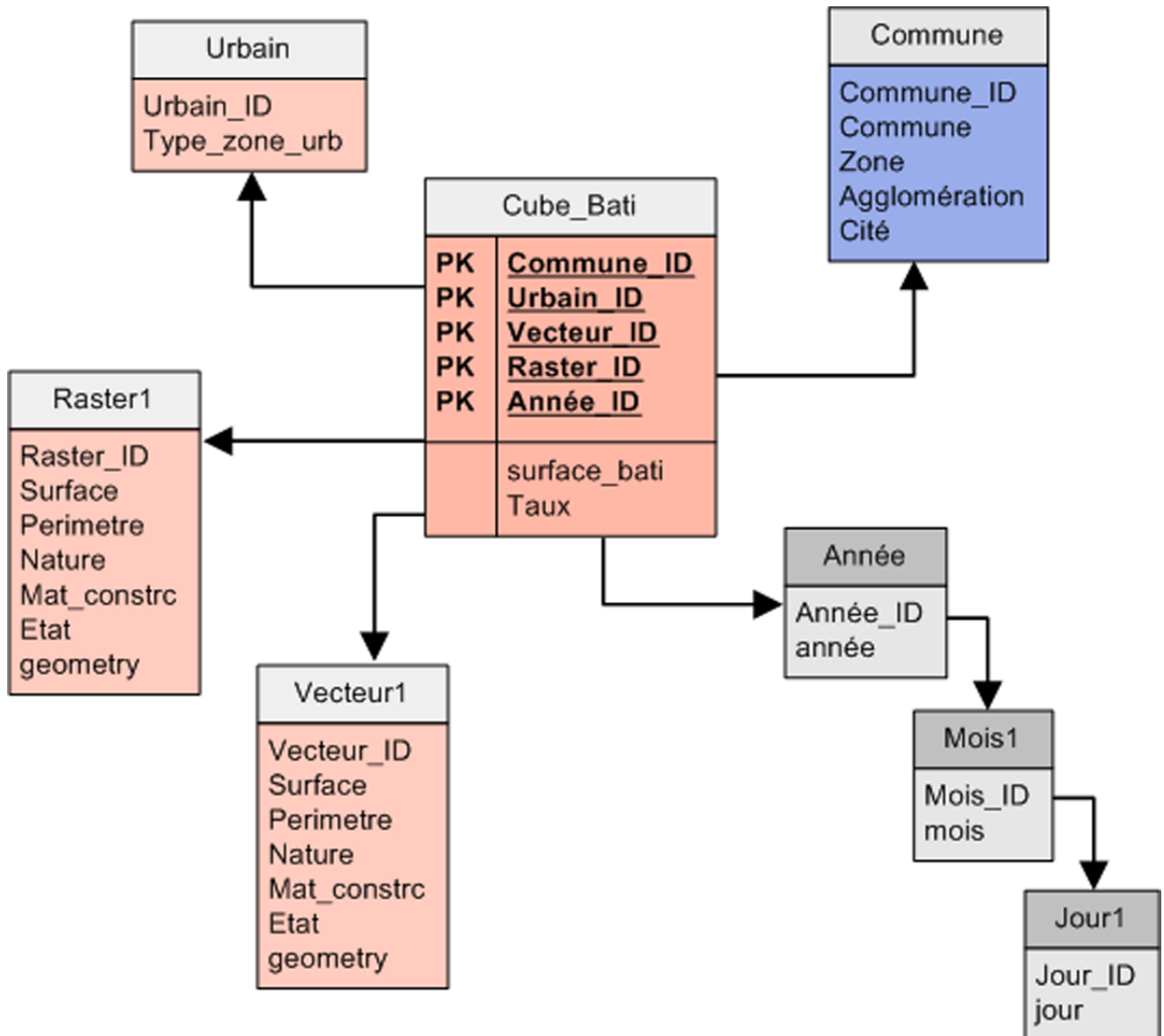


FIGURE 3.5 – Schéma en étoile du modèle bâti

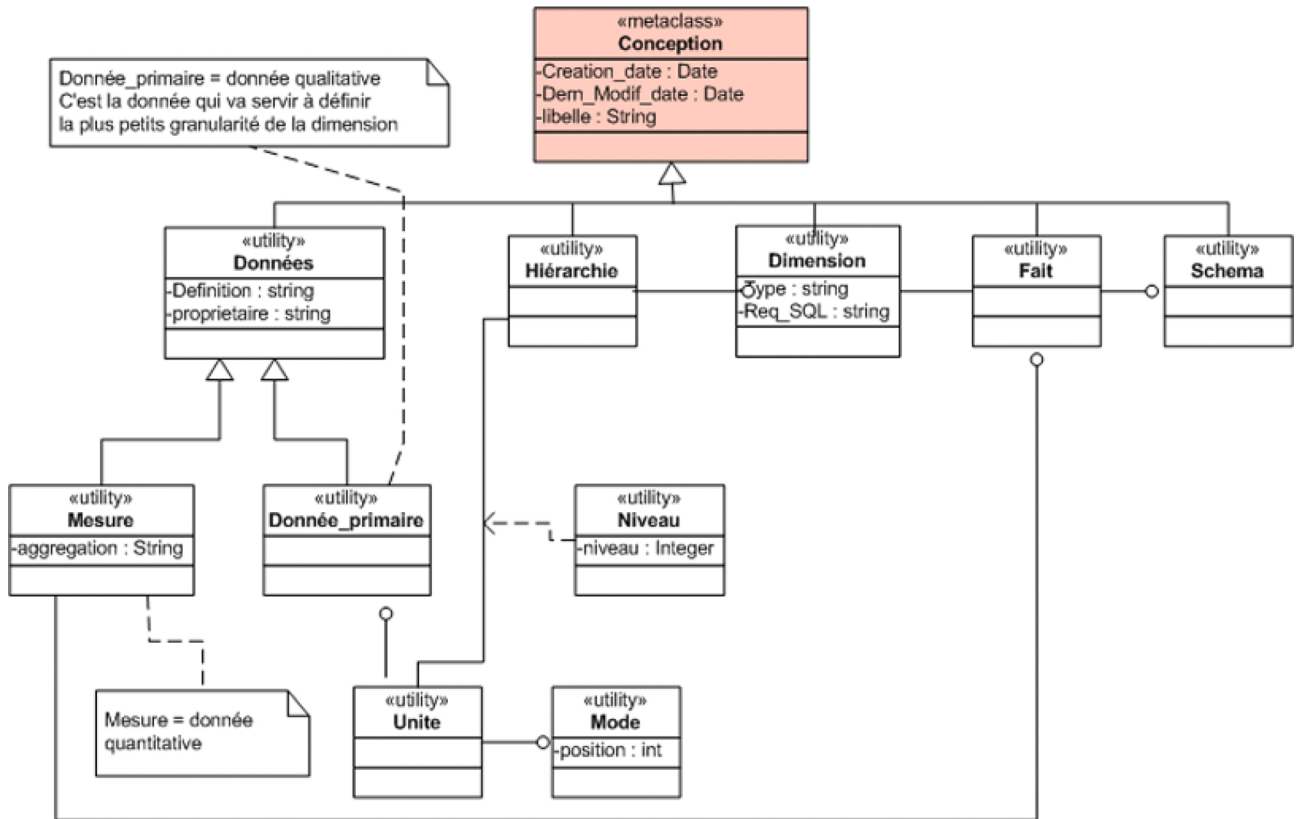


FIGURE 3.6 – Instance du processus d’implémentation de GESSlgh

En effet dans cette démarche plusieurs cas particuliers demandaient des traitements spéciaux, aussi il est au-delà de ce travail de développer un ETL adapté à notre système. Pour bien illustrer la gestion des métadonnées de l’entrepôt nous avons la représenter sous forme de diagramme de classe dans la figure 3.6. Nous envisageons que notre système soit complet, de ce fait les métadonnées présentées ici permettent de définir rapidement quelles sont les données qualitatives potentielles vouées à devenir des dimensions et quelles sont les données quantitatives vouées à devenir des mesures. Tout dépend de la table de la base des métadonnées soit elle regroupe toutes les unités avec ou sans notion de niveau (hiérarchie), puis les unités regroupent ou pas forcément les unités des niveaux inférieurs. Cette phase est réalisable automatiquement ou manuellement.

Une fois cette base de métadonnées renseignée, la création et le remplissage des tables de fait et des dimensions est lancé par la suite.

Pour terminer, il reste à charger les attributs géométriques des dimensions spatiales. Pour les fichiers Shape, il suffit d’exécuter le script *shp2pgsql* dans des tables vectorielles nommées de même nom que les fichiers Shape.

b) Chargement de la table de fait Nous avons pris le choix de la granularité au niveau du domaine d'étude (le bâti) et non pas au niveau des unités de la dimension "commune", donc les mesures ne sont pas forcément agrégées dans la table de fait ; Il n'est pas nécessaire d'effectuer des pré-calculs avant de charger les données. on peut calculer la mesure par la voie d'une requête spatiale (*Distance()*, *Intersects()*, *Contains()* et *Within()*) pour chaque zone la surface bâti, l'évolution du bâti et des variables. puis on peut agréger sur ces résultats et obtenir la somme et les moyennes par des niveaux hiérarchiques. Ces résultats agrégés ne sont plus qu'à charger dans les tables de faits.

De plus tous les résultats n'ont pas été chargés. c'est l'un des avantages des entrepôt de données est qu'il comporte exactement les données dont les utilisateurs ont besoins sous forme des magasins des données.

3.3 Bilan

Cette première partie de mise en place d'un système d'analyse spatio-temporelle sous Web à base d'entrepôt de données géographique se résume ainsi : La modélisation physique des dimensions qui vont stocker et structurer les données, l'élaboration des schémas en étoile (un modèle en étoile pour chaque domaine), le regroupement des différents schémas dans un seule modèle en constellation de fait, la mise en place des protocoles d'interrogation (requêtes) pour les processus métiers envisagées(seule le bâti a été pris comme cas d'étude) et enfin l'extraction et l'intégration des données spécialement l'intégration des informations géographique.

Un des premier objectif d'une structure décisionnelle au niveau de la commune de Laghouat est la production des résultats spatio-temporels. Cet objectif est atteint grâce à l'intégration de l'information géographique avec ses trois composantes vue dans le chapitre 2 plus l'aspect temporel relative à cette information que l'on a traité comme étant une quatrième composante.

Dans le prochain chapitre une implémentation d'un client Web-EDG offrant au utilisateurs la possibilité de faire une analyse spatio-temporel avec une exploration conviviale et rapide.

Chapitre 4

Étude expérimentale

Dans ce chapitre nous présentons notre prototype Web-EDG, c'est une solution SIG Web basée sur un modèle présenté dans le chapitre précédent qui repose sur les aspects OLAP-SIG classiques en rajoutant un support pour l'information géographique. Ce prototype fournit une interface intuitive et interactive, dont la conception a été guidé par le processus métier, afin visualiser les données de l'EDG de manière exploratoire.

Premièrement, nous décrirons l'architecture mise en place pour notre prototype. Nous montrons, comment elle s'appuie sur l'entrepôt de données géographique. Puis, nous présenterons notre réalisation en détail, ainsi que le fonctionnement de cette nouvelle approche. Enfin, nous conclurons ce chapitre par un bilan des caractéristiques obtenues et les limites, par rapport aux travaux connexes et en rapport avec les objectives visées.

4.1 Démarche d'implémentation du prototype Web-mapping à base d'EDG

Dans cette section nous commençons par introduire l'architecture de GESSlugh et la présentation des fonctionnalités SIG fournis par ce dernier. Ainsi que les justificatifs des différents choix d'implémentation fixés.

4.1.1 Architecture de la solution Web-EDG

La solution Web réalisée repose sur une architecture 3-tiers, comme montre la Figure 4.1, elle est composée :

- ◇ d'un Entrepôt de données géographique (orienté métiers en plusieurs magasins de données ex : le Bâti) dans lequel sont stockées les données géographiques et tabulaires PostgreSQL muni de la cartouche spatiale PostGis,
- ◇ d'un serveur cartographique GeoServer qui permet de traiter les requêtes et d'afficher les données géographiques,
- ◇ d'un client Web dans lequel l'utilisateur visualise les résultats des requêtes cartographiques et tabulaires. Ce dernier intègre un client SIG GeoExplorer, un plugin basé sur la librairie JavaScript OpenLayers.

4.1.1.1 L'entrepôt de données géographique

L'entrepôt de données géographique est implémenté en utilisant PostgreSQL¹ augmenté par sa cartouche spatiale PostGis². Les données géographiques et tabulaires sont stockées dans le SGBD PostgreSQL en utilisant la modélisation en étoile et en constellation de faits conçue précédemment (chap 4 § 2.1.2).

PostgreSQL appelé aussi Postgres, est un SGBD Objet-relationnel libre largement reconnu pour son comportement stable, proche de géant *Oracle* (propriétaire). Il fonctionne sur diverses plates-formes. Postgres peut être étendu pour supporter et stocker plus de types de données que les types traditionnels. En effet, il peut être étendu par un plugin appelé "PostGIS". C'est une extension (plugin) qui active la manipulation d'informations géographiques sous forme de géométries (points, lignes, polygones). PostGIS permet à PostgreSQL d'être un SGBD spatial.

1. La version "PostgreSQL 9.3.1"

2. La version "PostGis 2.1.1"

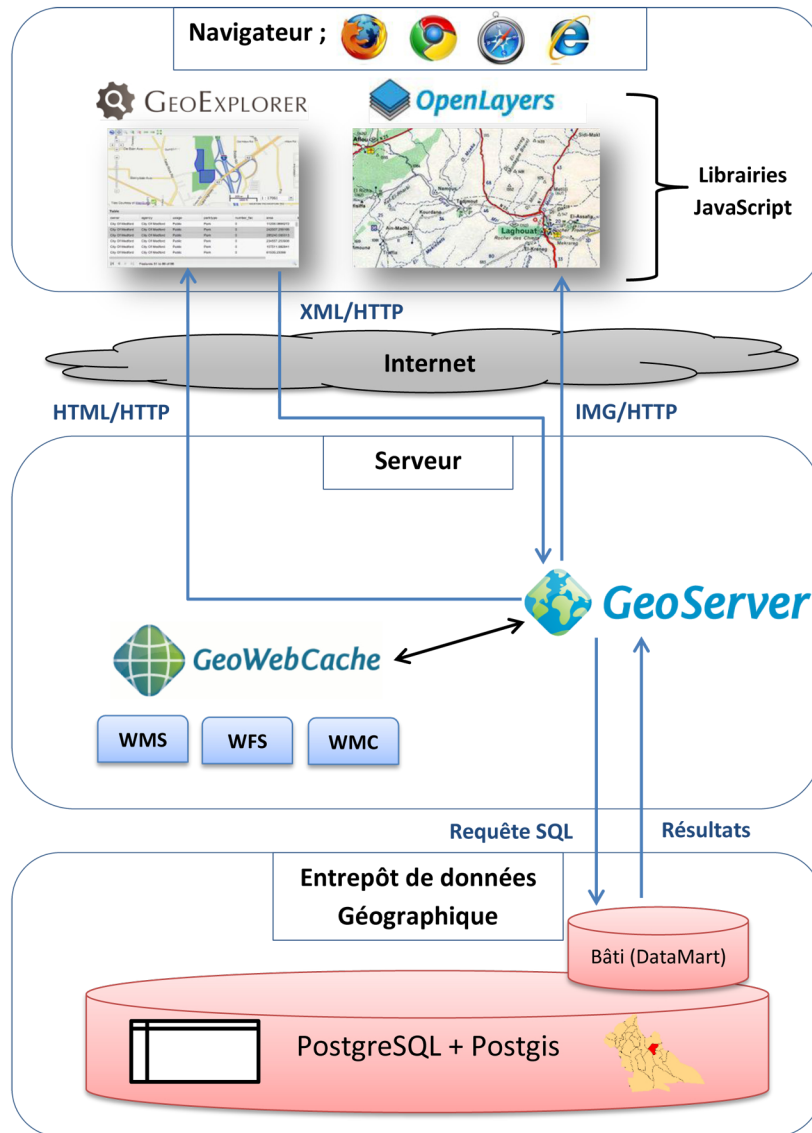


FIGURE 4.1 – Architecture de GESSlagh

Enfin, l'EDG mémorise aussi les informations concernant la dimension qui doit être cartographiée (les dimensions géographiques : Vecteur et Raster).

4.1.1.2 Le Serveur GeoServer

GeoServer est un logiciel Open-Source du domaine des logiciels libres qui permet de diffuser et de modifier des données géospatiales sur le Web. Il fonctionne comme un nœud dans une infrastructure de données spatiales libre et ouverte. Tout comme le serveur HTTP Apache qui offre une solution de serveur Web libre pour publier du HTML, GeoServer vise à faire de même pour les données géospatiales.

GeoServer est devenu l'implémentation de référence, standards de facto pour la diffusion de données selon les normes WFS et WCS (services de données vecteur et raster, respectivement), et reste l'un des meilleurs logiciels pour diffuser de cartes (assemblage de couches stylées) en WMS[Jé13].

GeoServer comportent plusieurs types de Webservices [Jé13] :

- ◇ WMS : Web Map Service, le plus connu, diffuse des cartes sous forme d'images raster, et renvoie optionnellement des données attributaires correspondant à un point (cliqué sur une interface par exemple).
- ◇ WMS-T : WMS Tuilé (dalles), plus rapide à utiliser car les tuiles sont mieux gérées par les caches (serveur et navigateur), par exemple : le GéoPortail IGN.
- ◇ WFS : Web Feature Service, diffusion d'entités géométriques vectorielles. Le format de ces entités est souvent le GML, mais on voit apparaître de plus en plus le GeoJSON.
- ◇ WFS-T : variante transactionnelle du précédent, c'est à dire qu'elle accepte les demandes de modification des données vectorielles (création / modification / suppression). Autorise la saisie ou la modification à distance (souvent avec une source PostGIS).
- ◇ WCS : Web Coverage Service, diffusion de données raster.
- ◇ WPS : Web Processing Service, exécution de traitements à distance. Par exemple on peut donner accès à des commandes d'un logiciel de SIG à distance, en fournissant les données et les paramètres du traitement à réaliser.
- ◇ CSW : Catalog Service for the Web, diffusion de catalogue de metadonnées.

4.1.1.3 Le client Web

Le client Web de notre application est développé à l'aide des deux composants logiciels :

1. La bibliothèque JavaScript *OpenLayers* <http://openlayers.org/>
2. et le framework *GeoExplorer* <http://suite.opengeo.org/docs/latest/geoexplorer/>

*OpenLayers*³ est une bibliothèque de fonctions JavaScript assurant un noyau de fonctionnalités orienté vers la mise en place d'applications clientes Web cartographiques fluides. *OpenLayers* permet d'afficher des fonds cartographiques tuilés ainsi que des marqueurs provenant d'une grande variété de sources de données. Une partie de cette bibliothèque permet aussi de gérer l'ergonomie proposée à l'utilisateur, elle 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, elle est intégrée dans l'API *GeoExplorer* ce dernier permet à l'utilisateur de créer des cartes thématiques, et d'introduire dans l'application Web cartographique de simples fonctionnalités d'analyse spatiale.

L'interface de notre prototype est baptisée *GESSlagh*, sera décrite dans la section suivante. Elle représente un environnement interactif d'aide à la décision qui encapsule la structure de l'application spatio-temporelle et traduit les interactions effectuées sur elle en opérateurs SIG multidimensionnels. Elle combine et synchronise la composante cartographique et l'aspect temporel de l'information géographique, la composante cartographique est une carte interactive qui répond visuellement aux actions de l'utilisateur et la composante temporelle est représentée sous forme d'un axe de temps qui permet à l'utilisateur de visualiser à travers le temps l'évolution du bâti.

4.1.2 Interface du système

Les Figures 5.2 à 5.7 représentent les captures d'écran les plus importantes de notre interface *GESSlagh*. En effet, cette interface Web fournit à l'utilisateur (Décideur, simple internaute) trois fonctionnalités principales sous les onglets :

3. La version *OpenLayers* 2.13.1



FIGURE 4.2 – La page principale de l’interface GESSlagh

La vue générale ① qui permet de visualiser et de naviguer dans l’ensemble des couches superposées,

La vue comparative ② permet de visualiser deux couches séparées,

L’analyse temporelle ③ permet la navigation dans l’historique de la zone visualisée.

L’interface est fédérée par une page d’accueil (Figure 4.2) où sont affichés certains images historiques de la villes ainsi que des informations sur l’outil GESSlagh.

4.1.2.1 La Vue générale

Cet onglet, illustré par la Figure 4.3, permet de regrouper toutes les composantes géographiques via les différentes couches dans un seul environnement. De plus, il présente un barre d’outils qui fournit l’accès aux différents opérateurs SIG.

Dans cette barre d’outils SIG, nous proposons des fonctionnalités purement SIG :

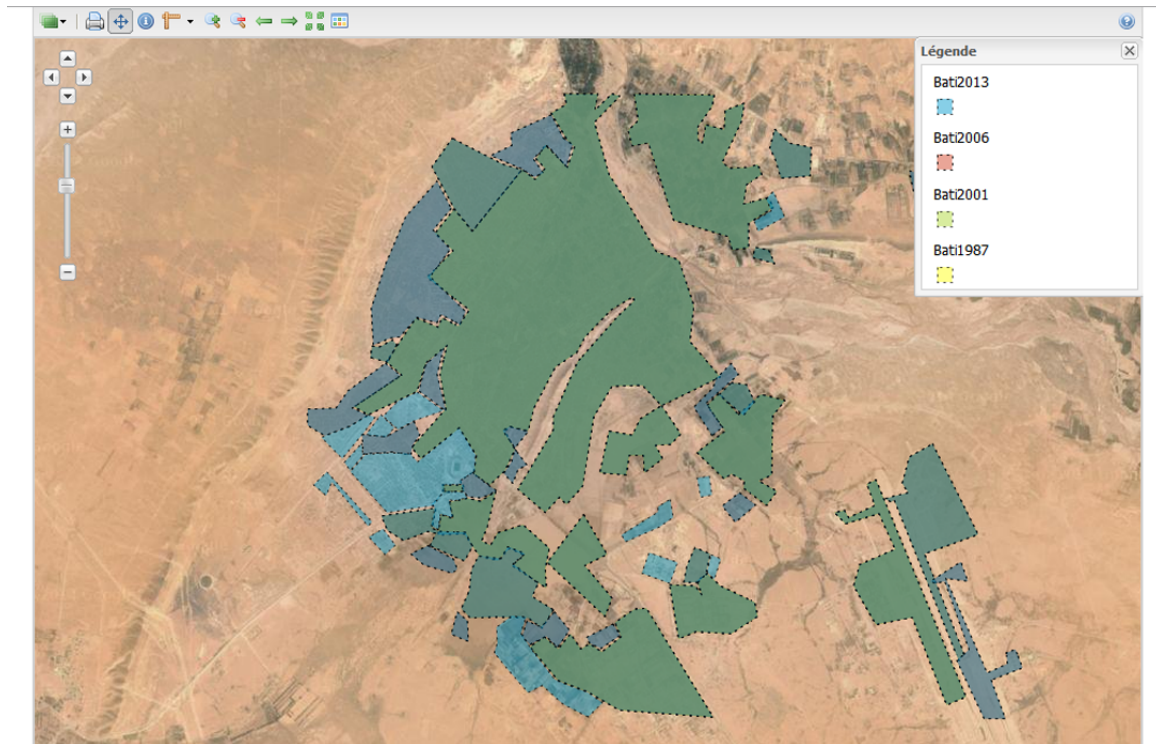


FIGURE 4.3 – L’onglet : Vue générale

1. *Zoom in* et *Zoom out* qui agrandissent et diminuent respectivement la taille de l’image de la carte
2. *Pan* qui permet de déplacer la carte
3. *Info Tool* qui permet d’avoir toutes les information alphanumériques associées aux objets géographiques (Figure 4.4)
4. *Map Print* qui permet d’imprimer ce qui afficher sur l’écran
5. *Rule Tool* qui calcule la distance entre deux points
6. le *Control Layer* (Figure 4.5). Ceci permet d’ajouter des couches vectorielles et/ou raster et de personnaliser l’affichage de la carte

Toutes ces fonctionnalités sont assurées par la manipulation de la bibliothèque OpenLayers. Cependant, parmi les codes développés, nous avons implémenté des fonctions JavaScript permettant d’extraire les informations tabulaires d’une couche vectorielle. Cette fonction "`GetFeaturInfos()`" est présentée dans l’annexe A.

Remarque : Un traitement particulier de la palette de couleurs superposées et nécessaire.

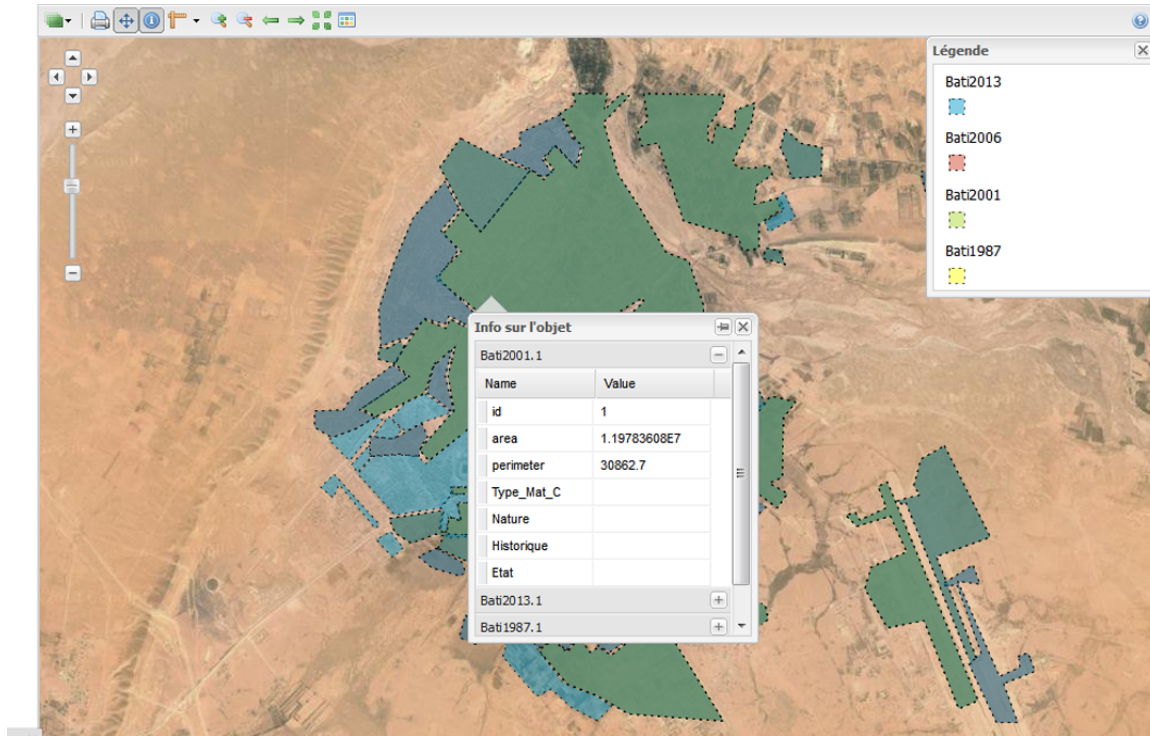


FIGURE 4.4 – L'onglet : Vue générale - Info Tool

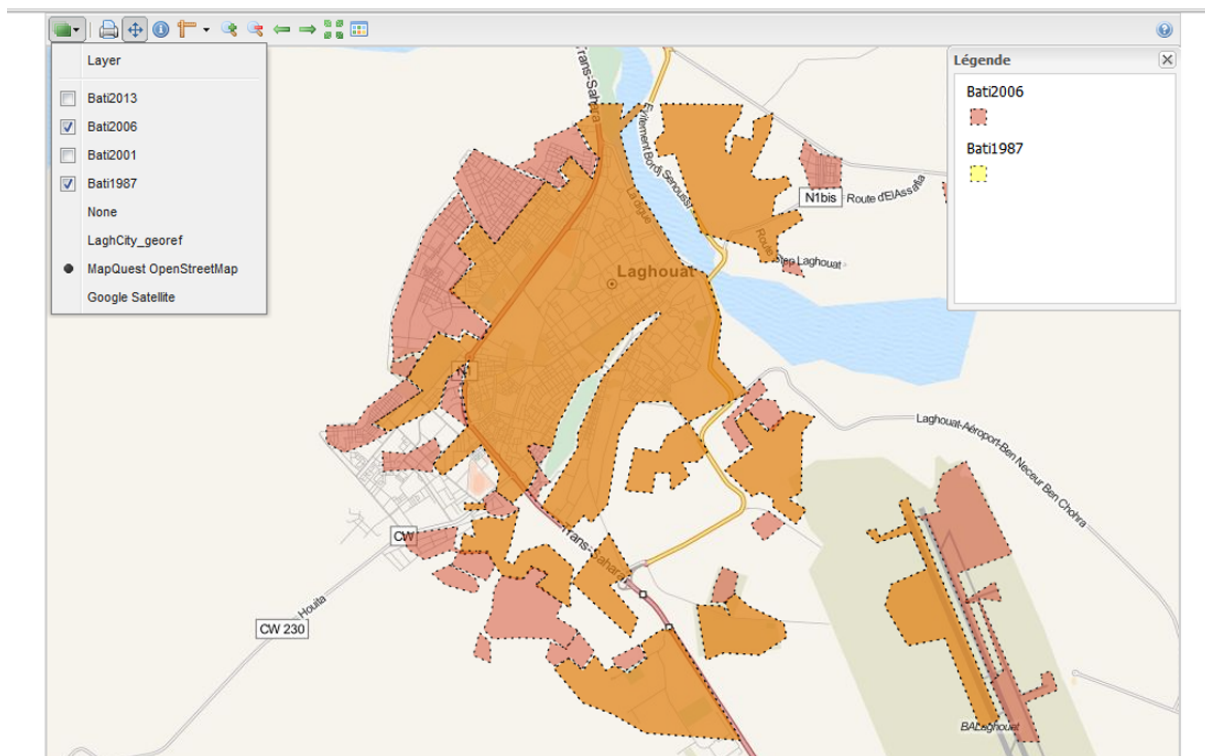


FIGURE 4.5 – L'onglet : Vue générale - Contrôle de couches

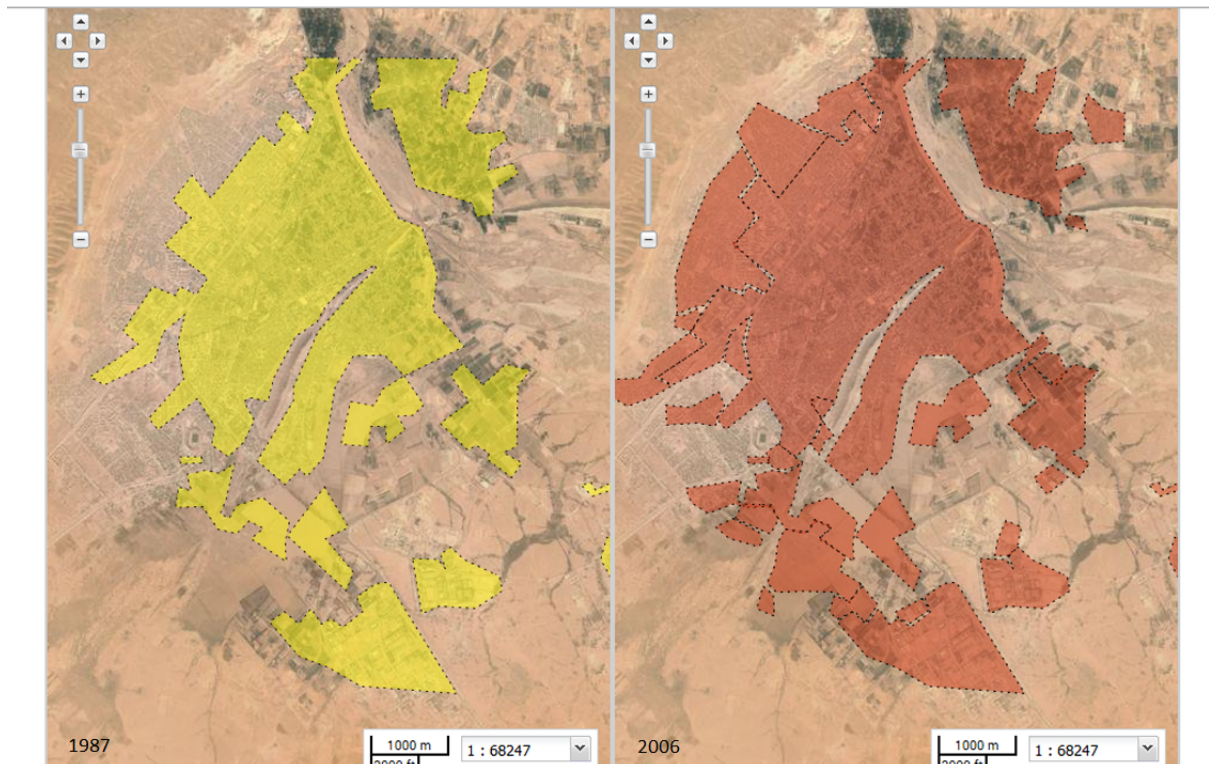


FIGURE 4.6 – L’onglet Vue Comparatif : Split

4.1.2.2 La Vue Comparative

Contrairement à la visualisation cartographique classique (superposition des couches), nous avons offert aux utilisateurs la possibilité de visualiser deux couches (trois, quatre respectivement) de la même zone géographique en même temps sur un split d’écran (ou ce qu’on appelle en anglais la visualisation Side by Side). En effet, ce besoin a été exprimé par les utilisateurs spécialisés dans les démarches d’identification des processus métiers (analyse des besoins).

La figure 4.6 illustre un exemple de cette technique de visualisation permettant une analyse spatiale plus conviviale. Sur le cadran gauche en vert pistache (couleur initiale : jaune) la couche qui représente le Bâti en 1987 et dans le cadran droit la couche en rouge brique qui représente le Bâti en 2006.

Nous soulignons qu’à cette étape, nous pouvons afficher toutes les informations tabulaires liées aux couches visualisées, telque (La surface, le périmètre, type de matériaux de construction, la nature d’usage,...etc).

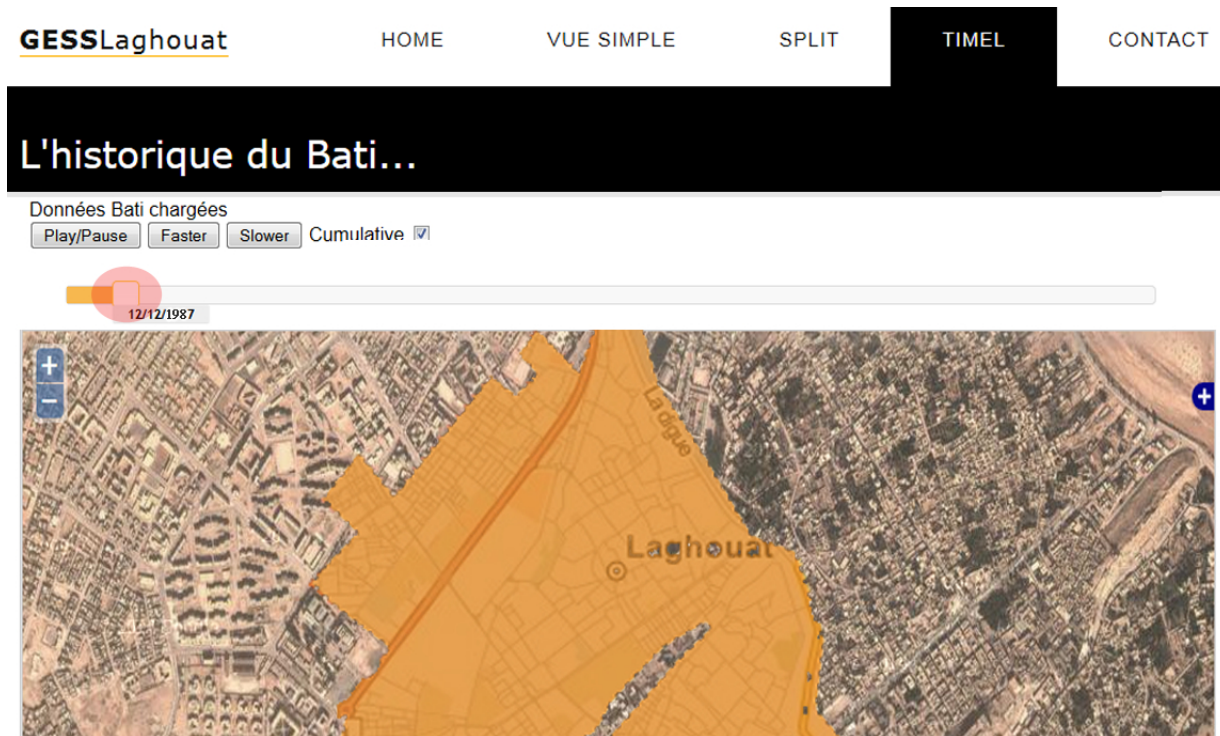


FIGURE 4.7 – L’onglet TimeL : Analyse temporelle

4.1.2.3 L’analyse temporelle

L’onglet TimeL, exprime une bonne partie de la puissance de l’outil GESSlagh. Il donne aux utilisateurs la possibilité de revenir dans le temps et visualiser les événements marquants sous forme de couches vectorielles. Par exemple, la réponse à la question : *"comment était l’état du Bâti dans l’année 1987 ?, et nous sommes en l’an 2014"*.

Dans la Figure 4.7, le bouton glissière (cercle rouge) permet à l’utilisateur de visualiser les couches en naviguant dans l’axe de temps (en couleur orange). L’utilisateur peut aussi lancer une navigation automatique sans devoir utiliser la glissière, cela en cliquant sur le bouton *Play/Pause*. Une option permet à l’utilisateur d’accélérer ou ralentir la navigation avec les boutons *Faster* et *Slower*.

4.1.3 Fonctionnalités

GESSlagh est une solution Web-SIG intégrée basée sur un EDG, un serveur cartographique et une interface qui représente un outil d’analyse spatio-temporelle (c-à-d : elle unifie un client OLAP et un client SIG).

GESSlagh, exploite les fonctionnalités du plugin *Timeline()* de la bibliothèque *jQuery-UI*.⁴

Ainsi, GESSlagh fournit les fonctionnalités avancées de l'analyse temporelle d'un serveur OLAP, tel que la hiérarchie temporelle automatique et les types de données standards pour la dimension temporelle, l'utilisateur peut se servir de l'axe du temps à partir des dates.

GESSlagh est basé sur le modèle multidimensionnel proposé dans le chapitre précédent. Il fournit une vision combinant les mesures et les dimensions, permettant de gérer outre les mesures classiques numériques, les mesures complexes géographiques et temporelles et aussi les dimensions géographiques et temporelles. En addition, des opérateurs de navigation multidimensionnelle de forrage, GESSlagh introduit deux nouveaux opérateurs qui adaptent au contexte spatio-temporel les opérateurs spatiaux de superposition et les opérateurs temporels de TimeLine.

L'interface Web de GESSlagh est conçue spécialement pour supporter l'analyse spatio-temporelle. Elle regroupe dans un seul environnement simple et interactif trois composantes différentes : cartographique, tabulaire et temporelle. Chaque interaction avec une de ces composantes est traduite en une opération de navigation multidimensionnelle ou SIG. L'objet au cœur de GESSlagh, est le composant logiciel *MapFrame()* (son code est présenté dans l'annexe A). Ce composant logiciel est né de la fusion du composant Map d'OpenLayers et de la balise *iFrame* de HTML, c'est à lui qui a la charge de construire le conteneur de la carte.

La puissance de GESSlagh réside dans l'onglet Analyse temporelle "TimeL", grâce au TimeLine l'utilisateur peut se lancer dans la visualisation de l'historique du Bâti. La Section 3 dans l'annexe A, présente le fragment du code de notre approche d'analyse spatio-temporelle "TimeLine".

4.2 Test de Performances

Comme tout système informatique, GESSlagh doit passer par une série de tests afin de prouver son efficacité et robustesse. En effet, nous devons s'assurer que GESSlagh satisfait

4. jQuery une librairie JavaScript v1.4.4 disponible <http://jquery.org/license>

Service	Capacité
Utilisation mémoire	50 - 256 MB
Version de la JVM	Sun Microsystems Inc.: 1.6.0_38 (Java HotSpot(TM) Client VM)
Polices disponibles	GeoServer a accès à 512 polices de caractères différentes
Disponibilité du JAI ImageIO	TRUE
Mémoire maximale pour le JAI	247 MB
Nombre de tâches de tuilage JAI	7
Délai de maintien en vie du pool de threads ThreadPoolExecutor (ms)	30000
Séquence de mise-à-jour	175
Cache	500 MB (extensible)
Taille du buffer permettant de logger le contenu XML	1024 ko

TABLE 4.1 – Résumé de la configuration du GeoServer

les requêtes des utilisateurs avec des résultats fiables et dans des délais raisonnables ; soit un temps d'attente le plus petit possible.

Concernant les temps de réponses du système, notre stratégie de test était de simuler un fonctionnement sur une machine avec une connexion ADSL de **128 ko/s**. Du côté client, l'application GESSlagh ne nécessite pas une configuration particulière vu que notre système est déployé via le Web. Du côté serveur, une machine de configuration standard doté d'un serveur cartographique (GeoServer) ayant les caractéristiques de base illustrées dans le tableau 4.1.

4.2.1 Satisfaction des utilisateurs

Concernant la fiabilité des résultats et la satisfaction des requêtes des utilisateurs, GESSlagh reste un prototype non complet vu le manque de données et l'absence des professionnels des métiers (telque les experts du domaine de Bâti). En effet, l'équipe du SIG est nouvellement créée au sein du laboratoire d'informatique et de mathématique (LIM) à l'université de Laghouat, et très peu de relation de coopération université-entreprise sont

instaurées.

Cependant, nous avons eu un seul feedback des chercheurs en télédétection et gestion des territoires. Malgré que ce feedback est positif, nous nous pouvons tirer des conclusions fermes.

4.2.2 Délais d'attente

Vis a vis des trois principales fonctionnalités de notre système, nous avons mesuré les délais d'attente de la satisfaction des requêtes.

L'interaction avec le site Web-mapping fait appel à plusieurs fichiers de tailles différentes. Ces fichiers peuvent être soit HTML, JS, CSS, Jpeg, Png ou Gif.

Dans un premier temps, nous avons effectué nos premiers tests en mode *LocalHost*, dans le but de recenser tous le travail en arrière plan du système, étant composé de plusieurs outils. La vitesse de lecture depuis le cache du serveur locale est de : **680 Ko/s**. Dans ce mode LocalHost, nous utilisons Internet *ADSL* avec un débit de : **128 Ko/s** juste pour visualiser la carte de fond (source : Google Satellite).

Pour toutes les fonctionnalités du système, l'ensemble des requêtes cartographiques classiques que nous avons appelées tâches en arrière plan : chargement de carte depuis le cache, identification de la résolution, niveau de zoom,...etc. La taille est mesurée en *Kilo Octets* et la vitesse en *Kilo Octets / Seconde*.

Les Figures 4.8, 4.9 et 4.10 illustrent les résultats des requêtes de tests effectués sur les trois fonctionnalités respectivement.

Ces figures donnent des détails trop amples qui nous permettent de voir tous ce qui se fait en arrière plan, les fichiers chargés, les codes exécutés, les cartes visualisées, ainsi le temps consommé.

Dans le Tableau 4.2, nous résumons les résultats obtenues après les tests effectuées selon les 3 fonctionnalités de GESSlagh.

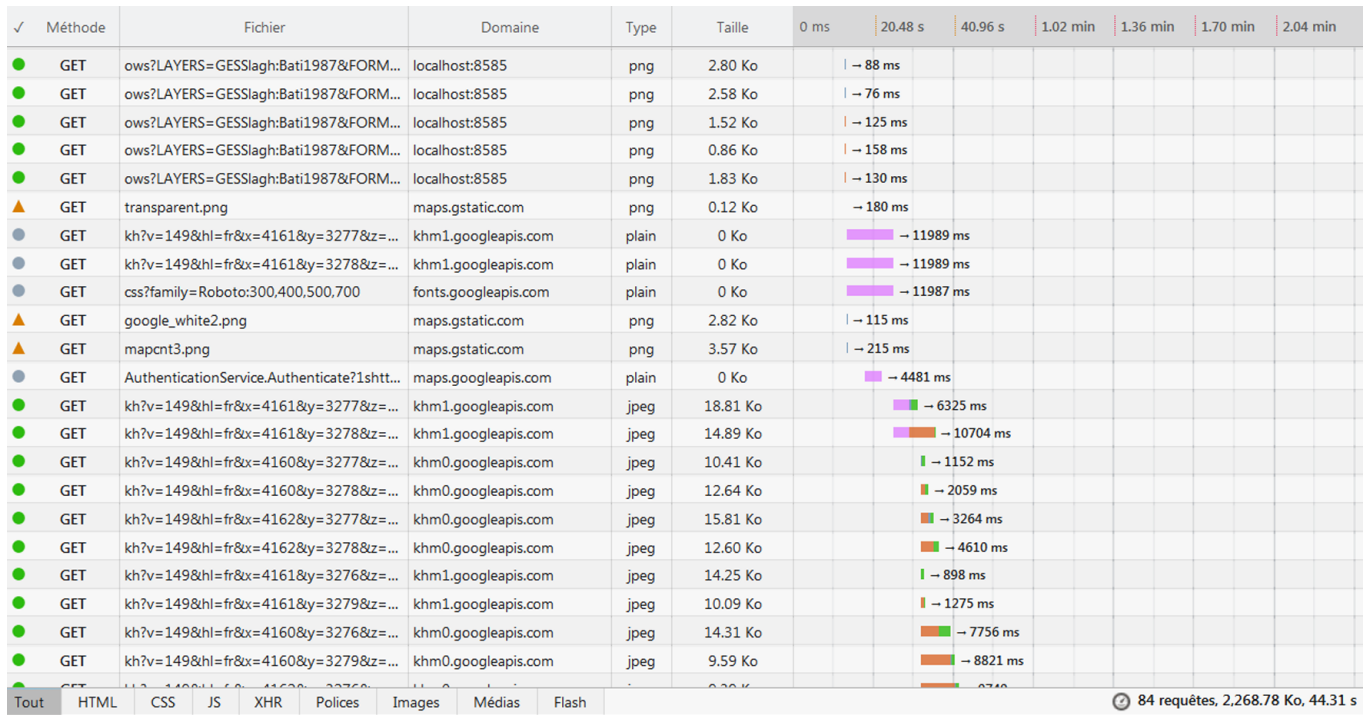


FIGURE 4.8 – Tests de performances sur l’onglet Vue simple

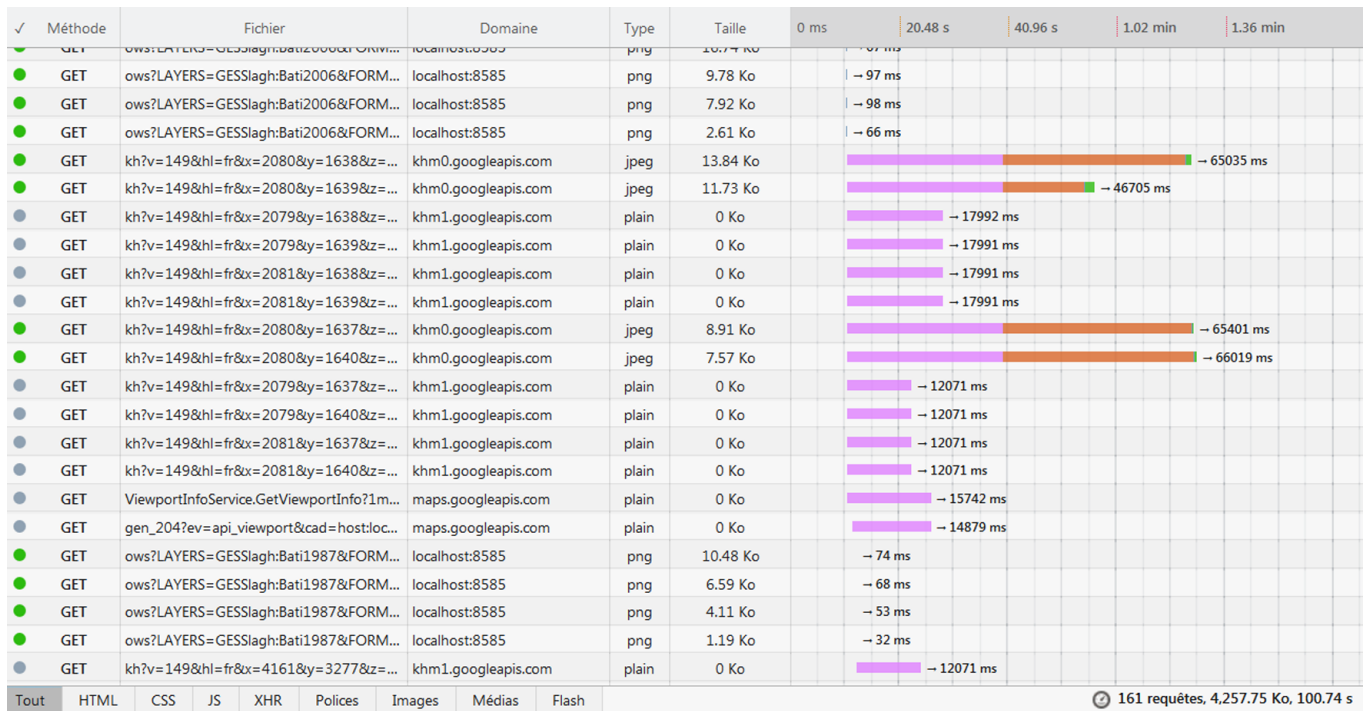


FIGURE 4.9 – Tests de performances sur l’onglet Vue comparative

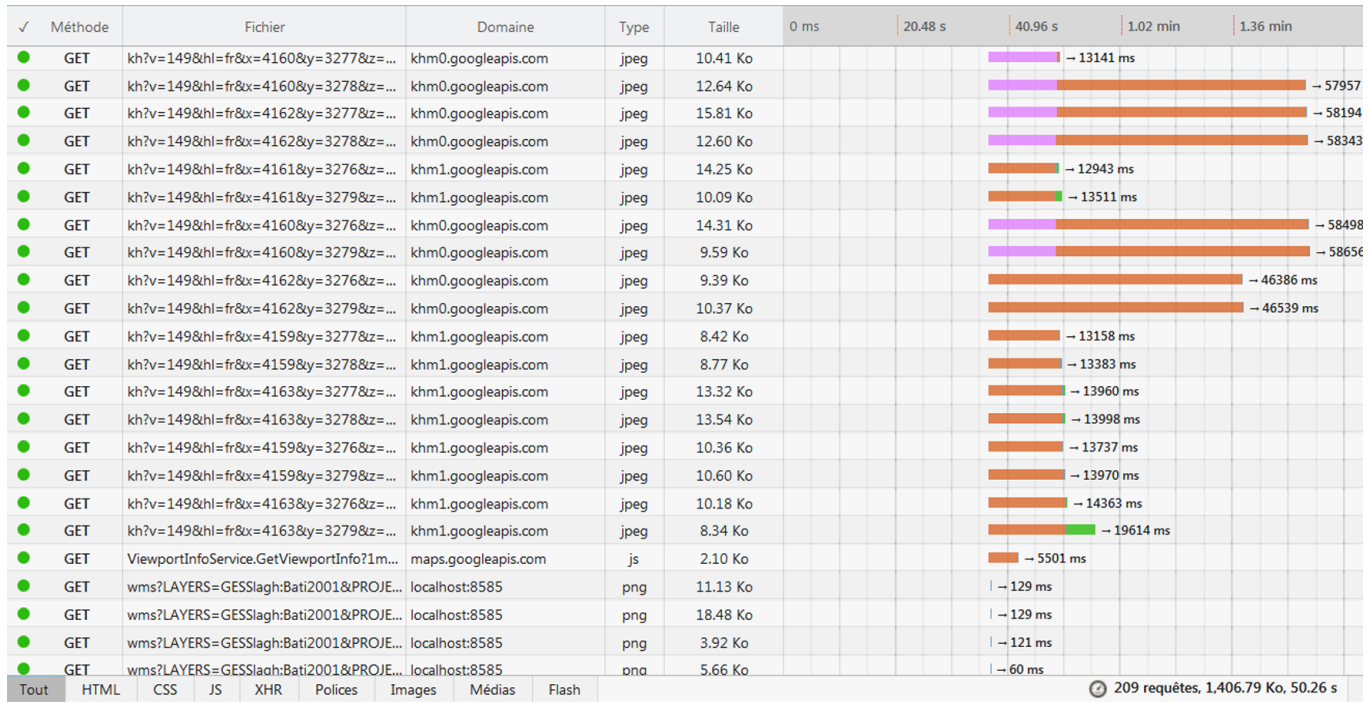


FIGURE 4.10 – Tests de performance sur l’onglet TimeLine

Onglet	JavaScript			Images			HTML / CSS		
	Nbr	Taille	Temps	Nbr	Taille	Temps	Nbr	Taille	Temps
Vue Simple	3	1,714.65	2.18	34	82.01	0.65	12	212.59	1.75
Vue Comparative	19	3,708.33	23.94	70	325.33	32.37	26	427.38	24.58
Analyse Temporelle	9	298.40	15.77	43	559.83	21.81	5	1.01	0.63

Nbr : nombre de requêtes La Taille en Ko Le Temps en Secondes

TABLE 4.2 – Résumé des résultats de test en terme de tâches réelles effectuées

Vu que ces données expérimentales ne sont pas réalistes car l’on travaille en mode LocalHost, nous avons fait recours à une estimation proche de la réalité. En effet, nous avons pris comme modèle réseau une connexion ADSL à 1MB/s (soit 128 Ko/s vitesse

réel de téléchargement).

Nous avons effectué 15 requêtes distribués comme suit :

- ◇ 5 requêtes sur la fonctionnalité Vue simple ($Vs1, Vs2, \dots, Vs5$),
- ◇ 5 requêtes sur la fonctionnalité Vue comparative ($Vc1, Vc2, \dots, Vc5$),
- ◇ et 5 requêtes sur la fonctionnalité Analyse temporelle ($At1, At2, \dots, At5$).

Ce qui à généré un ensemble de 60 tâches en arrière plan :

- ◇ 5 tâches JavaScript de taille de 512ko,
- ◇ 50 tâches d'Images de taille moyenne de 512ko,
- ◇ et 5 tâches HTML/CSS de taille de 128ko.

Nous avons choisi des requêtes faisant appel à des vues simples d'une parcelle précise, avec des opérations d'agrandissements/réductions fixés et des requêtes de vue globale d'une région. Concernant les requêtes comparatives là aussi on a essayé de faire des manipulations diverses. En ce qui concerne les vues faisant appel à l'analyse temporelle nous avons fait des manipulations faisant appel à plusieurs dates.

Les trois courbes de la Figure 4.11 représentent les résultats obtenus pour chaque fonctionnalité de notre système en terme de temps de réponse par rapport au type de fichier chargé. Tandis que, les courbes de la Figure 4.12 représentent le temps d'attente de chaque requête des 3 fonctionnalités du système.

Interprétation

D'après les résultats de la figure 4.12, nous constatons que le délai maximum d'attente est de :

- ◇ 16 secondes pour la visualisation simple,
- ◇ 38.7 secondes pour la vue comparative,
- ◇ et 139.8 secondes pour l'analyse temporelle.

Ces délais peuvent être considérés largement raisonnables pour un analyste qui cherche à faire des comparaisons ou de l'analyse spatio-temporelle, chose qui se fait généralement en amont à une décision ou une action.

Le délai d'attente moyen par requête est de :

- ◇ 11 secondes pour la vue simple
- ◇ 41 secondes pour la vue comparative
- ◇ 130 secondes pour l'analyse temporelle

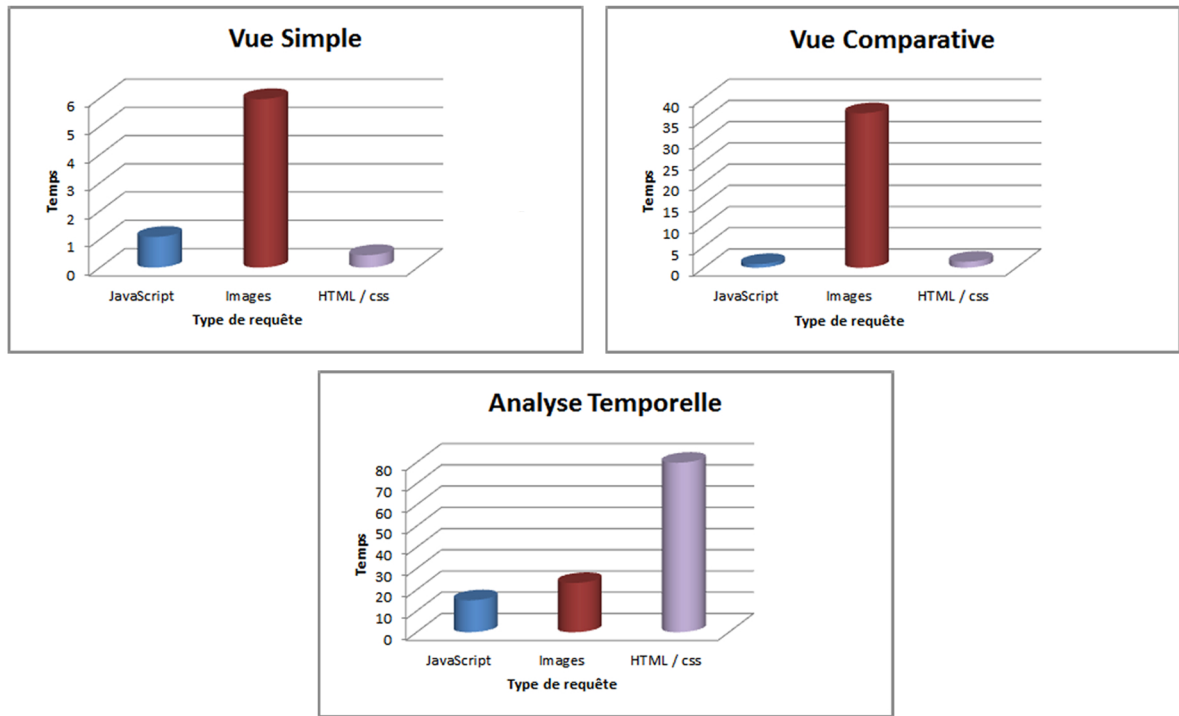


FIGURE 4.11 – Tests de performances de GESSlgh sur un réseau proche de la réalité - tâches en arrière plan

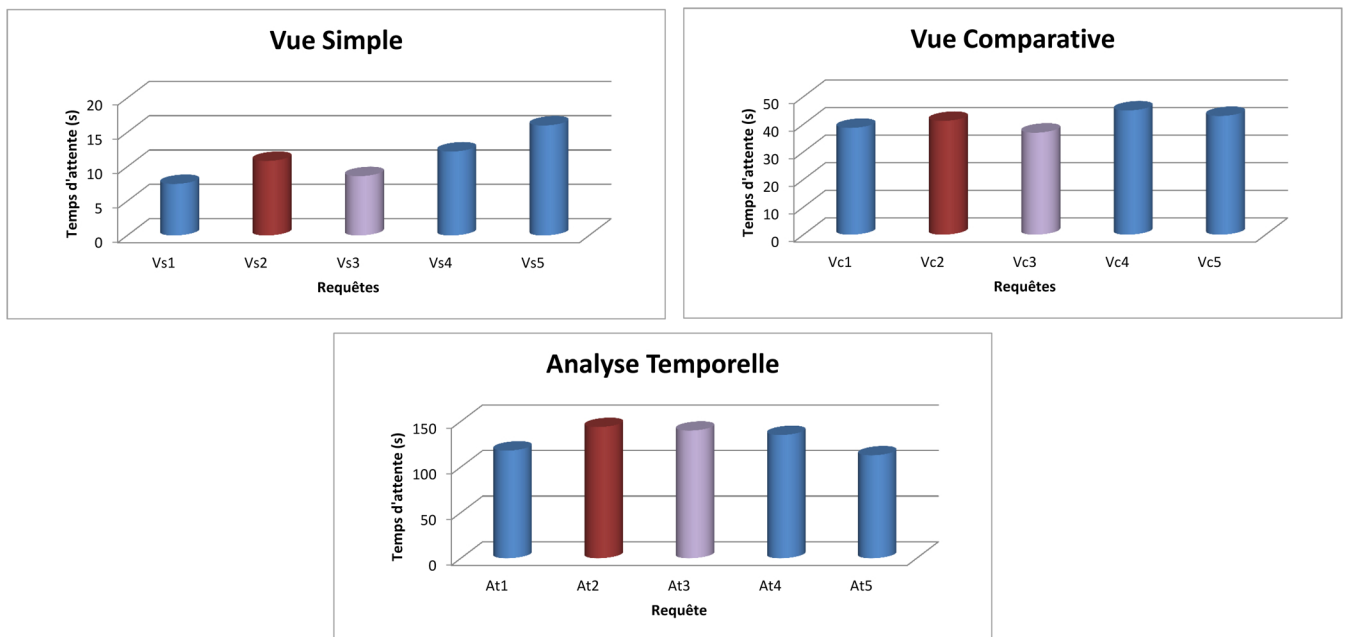


FIGURE 4.12 – Tests de performances de GESSlgh sur un réseau proche de la réalité.

Le délai d'attente moyen pour n'importe quel type de requête via l'application est de : 61 secondes

Mentionnons que ces serveurs cartographiques sont généralement mis dans le territoire national (au niveau de ministère tutelle, avec des sites miroirs au niveau de chaque wilaya), et où la connexion est plus fluide est moins contrainte.

4.3 Analyse comparative

Le système GESSlagh, basé sur une modélisation multidimensionnelle détaillée dans le chapitre précédent, présente deux avantages fondamentaux par rapport aux solutions des travaux présentées dans le chapitre 2 :

- ◇ La séparation des données géographiques (vecteur et raster) l'or de l'intégration de la composante spatiale,
- ◇ L'ajout d'une nouvelle composante "*La composante temporelle*" à l'information géographique.

En effet (Brisebois, 2002), (Pestana et al., 2005) et (Bimonte et al., 2006) gèrent seulement la composante spatiale de l'information géographique avec aucun détail sur les techniques d'intégration de cette dernière dans le support de stockage (BDD spatiale ou EDS), tandis que (Bimonte, 2008) intègre ainsi la composante sémantique.

En comparant la conception et les fonctionnalités de GESSlagh au travail proposé par (Selmoune et al., 2012), on trouve des similarités dans l'architecture globale mais avec très peu de détails sur la façon dont ils ont conçu et implémenté cette intégration de l'information géographique dans l'EDG, nous nous pouvons tirer des conclusions.

Par ailleurs, nous nous pouvons réaliser des comparaisons expérimentales, du fait que nous n'avons pas trouver un système qui travaille sur le même domaine d'étude à savoir l'évolution du bâti à travers le temps.

4.4 Limites

GESSlagh offre de nouvelles fonctionnalités Web mapping et il intègre la composante temporelle de l'information géographique. Cependant, il représente un noyau préliminaire à faire évoluer. Il présente différentes limites liées au manque d'implémentation de

certaines opérateurs de navigation dits "opérateurs OLAP", et l'implémentation d'autres fonctionnalités de visualisation.

Concernant les opérateurs OLAP, nous spécifions les opérateurs cartographiques seulement. Car du côté serveur OLAP, il offre plusieurs opérateurs d'analyse multidimensionnelle tel que l'opérateur Dice, Slice, l'opérateur de coupe,...etc.

Du point de vue visualisation, GESSlugh ne permet pas pour le moment de gérer la visualisation multidimensionnelle sous forme de cube, qui est une caractéristique importante des systèmes d'exploration et d'analyse des données à faire implémenter et exploiter dans nos travaux futurs.

4.5 Conclusion

Nous avons présenté dans ce chapitre notre prototype GESSlugh un outil Web-SIG basé sur un Entrepôt de données géographique. GESSlugh gère des données géographiques et complexes, et traite les dimensions géographiques séparant les données vecteur et raster.

Notre outil implémente trois fonctionnalités essentielles : *i)* visualisation simple de couches vectorielle/raster, *ii)* une visualisation comparative pour affronter des cartes pour les comparer et *iii)* une analyse spatio-temporelle où nous visualisons l'évolution du bâti à travers le temps.

Nous avons effectué une étude pour mesurer les performances de notre outil. Et nous avons terminé par une analyse comparative en terme de modélisation avec les travaux connexes.

Les limites de notre outil ont été aussi mentionnées.

Conclusion Générale

Vers la fin des années 1990 et pendant la révolution de l'informatique décisionnelle qui apporte des nouvelles solutions pour la modélisation, l'interrogation et la visualisation des données spatiales dans un objectif d'aide à la décision, les développeurs des SIG se sont orienté vers deux axes ; la représentation des données spatiales dans le Web "Web-mapping" et l'utilisation des données spatiales dans le processus d'aide à la décision "Géo-décisionnel". La fusion entre ces deux domaines donne un système "Web-SIG à base d'EDG", cela est bien plus qu'un mot à la mode puisqu'il ouvre un nouveau axe de recherche semble déjà très prometteur.

Dans ce contexte, notre travail a fourni un noyau d'une application Web-EDG qui permet de toucher de près des données spatio-géographiques.

En résumé, nos travaux concernant l'intégration de l'information géographique dans les entrepôts de données s'articulent autour de trois contributions.

- 1) Acquisition des savoirs-faire :** La première phase a été consacré principalement à appréhender les techniques liées aux entrepôts de données géographiques et au Web-mapping et à les mettre en œuvre. Le choix de l'architecture, ainsi que des logiciels et outils Open Source ont été décidés systématiquement après la phase de recherche et d'étude des travaux connexes, cette phase faisait un état des lieux des meilleurs solutions. La solution GeoServer/OpenLayers correspondait exactement aux besoins.
- 2) Le modèle conceptuel de données multidimensionnels :** L'introduction d'information géographique au entrepôt de données amène à une reformulation des concepts de ce dernier. Cela soulève un ensemble de problématiques de modélisation multidimensionnelle particulière. Nous proposons ainsi une adaptation du modèle multidimensionnel pour ne pas dire un nouveau modèle. Dans notre modèle on a séparé les données raster du vecteur en chargeant chaque type d'information géographique

avec une procédure de chargement différente. ainsi notre modèle traite l'aspect temporel comme une composante de l'information géographique.

3) Le prototype GESSlagh : Nous avons développé un logiciel Web-SIG alimenté par un entrepôt de données géographique. GESSlagh fondé sur notre modélisation multidimensionnelle proposé dans le chapitre 3, notre outil conçu pour satisfaire plus les requêtes décisionnelles, il présente les caractéristiques fondamentales d'un système d'analyse spatio-multidimensionnelle. De plus des opérateurs de navigation multidimensionnelle de forage, GESSlagh introduit deux nouveaux opérateurs qui adaptent au contexte spatio-temporel les opérateurs spatiaux de superposition et temporel de l'analyse temporelle.

Perspectives

Malgré ses performances, GESSlagh présente des limites du point de vue de la visualisation/interaction et de l'implémentation des opérateurs définis par l'OLAP Géographique.

GESSlagh, représente une première tentative d'automatisation d'un système OLAP à base d'EDG pour notre wilaya via le Web. Cependant, comme tout travail en ses débuts, notre apport est a amélioré. Comme perspectives, nous prévoyons plusieurs améliorations parmi les plus importantes :

Sur le plan conceptuel, nous envisageons mettre en place un méta-modèle pour faciliter la conception des EDG avec un processus de validation robuste comme pour les ED classique [W.H96] [Ral96].

Sur le plan implémentation, nous prévoyons améliorer l'outil de visualisation spatio-temporelle par l'ajout de nouvelles fonctionnalités d'analyse multidimensionnelle.

Acronymes

Cet liste représente l'ensemble des acronymes (ainsi que leurs significations) les plus utilisés dans ce mémoire.

ANRH Agence Algérienne des Ressources Hydraulique

BDD Base de Données

BDS Base de Données Spatiale

DW Data Warehouse

ED Entrepôt de Données

EDG Entrepôt de Données Géographique

EDS Entrepôt de Données Spatiale

ETL = **ETC** Extract-Transform-Load

FTP File Transfer Protocol

GESSLagh Gestion Socio-Spatiale ; Laghouat

HOLAP Hybrid On-Line Analytical Processing

JAI Java Advanced Imaging

MOLAP Multidimensionnel On-Line Analytical ProcessingacroJAIJava Advanced Imaging

OLAP On-Line Analytical Processing

OLTP On-Line Transactional Processing

ONS Office Nationale des Statistiques

PHP Hypertext Preprocessor

PVL Plug-ins for Visual Languages

ROLAP Relational On-Line Analytical Processing

SGBD Système de Gestion des Base de Données

SIG Systèmes d'Information Géographique

SIGHout Systèmes d'Information Géographique pour la wilaya de Laghouat

SOLAP Spatial On-Line Analytical Processing

TCP/IP Protocole d'Internet

UML Unified Modeling Language

WCS Web Coverage Service

Web-EDG Entrepôt de Données Géographique sous Web

WFS Web Feature Service

WMS Web Mapping Service

WPS Web Processing Service

Annexe A

.1 GetFeaturInfo()

Les bibliothèques d'OpenLayers offre des fonctionnalités SIG classiques. Nous avons ajouter une fonction qui permet d'extraire les attributs tabulaire liée au carte vectorielle. Son code est présenter ci-dessous :

```
map.events.register('click', map, function (e) {
  document.getElementById('nodelist').innerHTML = "Loading... please wait...";
  var params = {
    REQUEST: "GetFeatureInfo",
    EXCEPTIONS: "application/vnd.ogc.se_xml",
    BBOX: map.getExtent().toBBOX(),
    X: e.xy.x,
    Y: e.xy.y,
    INFO_FORMAT: 'text/html',
    QUERY_LAYERS: map.layers[0].params.LAYERS,
    FEATURE_COUNT: 50,
    Layers: 'GESSlagh:Bati1987, GESSlagh:Bati2001',
    Styles: '',
    Srs: 'EPSG:4326',
    WIDTH: map.size.w,
    HEIGHT: map.size.h,
    format: "image/jpeg" };
  OpenLayers.loadURL("http://localhost:8585/geoserver/GESSlagh/wms", params, this, setHTML, setHTML);
  OpenLayers.Event.stop(e);
});
```

.2 MapFrame

Appelé aussi (Map Container), le conteneur de la carte c'est une fusion du composant "Map" d'OpenLayers et la balise "iFrame" du HTML. Voici le code de MapFrame :

```
function init(){
  map = new OpenLayers.Map('map', {
    projection: new OpenLayers.Projection("EPSG:900913"),
    units: "meters",
    maxResolution: "auto",
    maxExtent: new OpenLayers.Bounds(-20037508.34,-20037508.34,20037508.34,
    20037508.34),
    zoom:15
  });

  var gsat = new OpenLayers.Layer.Google("Google Satellite", {
    type: google.maps.MapTypeId.SATELLITE,
    numZoomLevels: 16
  });

  map.addLayers([gsat]);
  map.addControl(new OpenLayers.Control.LayerSwitcher());
  map.zoomToExtent(new OpenLayers.Bounds(316316.4486908615,3995720.2197048906,
    326422.5424194146,4005993.723835196))
};
```

.3 Le TimeLine

```

OpenLayers.Timeline = OpenLayers.Class({
  map: undefined,
  slider: undefined,
  current_date: new Date().getTime() / 1000,
  display_layer: undefined,
  current_data: undefined,
  timerId: undefined,
  selectControl: undefined,
  selectedFeature: undefined,
  interval: 5,
  speeds: ["Really slow", "Slow", "Normal", "Fast", "Really fast"],
  curr_speed: undefined,
  data_format: undefined,
  cumulative: true,
  timedelta: 3600,
  first: undefined,
  onFeatureInsert: undefined,
  name_key: undefined,

  initialize: function(options) {
    this.map = options.map;
    this.display_layer = this.createDisplayLayer(),
    this.Bati1987_layer = this.createBati1987Layer(),
    this.Bati2001_layer = this.createBati2001Layer(),
    this.Bati2006_layer = this.createBati2006Layer(),
    this.Bati2013_layer = this.createBati2013Layer(),
    this.map.addLayer(this.display_layer);
    this.createSelectControl();
    this.slider = options.timeline;
    this.curr_speed = parseInt(this.speeds.length / 2);
    this.data_format = new options.format(options.format_options);
    this.name_key = options.name_key;
    if (options.date_key) {
      this.data_format.date_key = options.date_key;
    }
    if( options.timedelta) {
      this.timedelta = options.timedelta;
    }
    this.data_format.timestamp_func = options.date_func;
    this.onFeatureInsert = options.onFeatureInsert;

    var self = this;
    $(this.slider).slider({
      range: "min",
      min: 0,
      max: 100,
      value: 0,
      disabled: true,
      change: function (e, ui) {
        if (self.display_layer) {

          if (ui) {
            self.data_format.past_seconds = Math.ceil(self.first+(self.
              current_date-self.first)*(ui.value / 100.0));
          }
          if (!self.cumulative) {
            self.data_format.lowerlimit = self.data_format.past_seconds - self.

```

```

        timedelta;
    }
    else {
        self.data_format.lowerlimit = undefined;
    }

    if (!self.data_format.past_seconds) {
        self.data_format.past_seconds = 0; // first run
    }
    self.display_layer = self.updateDisplayLayer();

    self.createSelectControl();
}

var bounds = self.display_layer.getDataExtent();
if (bounds && self.cumulative) {
    self.map.zoomToExtent(bounds, false);
}
},
slide: function(e, ui) {
    self.stopBar();
    if (self.current_data) {
        self.data_format.past_seconds =
            Math.ceil(self.first+(self.current_date-self.first)*(ui.value / 100.0));
    }

    var d = new Date(self.data_format.past_seconds*1000);
    d = d.getDate()+"/"+(d.getMonth()+1)+"/"+d.getFullYear();
    var tooltip = $("

page 73


```

.4 Le dictionnaire de données

Le dictionnaire de données associe au Modèle en constellation de faits envisagé

<i>urbain</i>			
Champ	Type	Description	Remarque
URBAIN_ID	int	Numéro identifiant l'urbain	Clé primaire
ZONE_URB	text	le nom de la zone urbaine	
ZONE_AGRC	text	le nom de la zone agricole	
<i>BATI</i>			
Champ	Type	Description	Remarque
COMMUNE ID	int	Numéro identifiant la commune	Clé étrangère
URBAIN ID	int	Numéro identifiant l'urbain	Clé étrangère
VECTEUR ID	int	Numéro identifiant le vecteur	Clé étrangère
RASTER ID	int	Numéro identifiant le raster	Clé étrangère
ANNEE ID	int	Numéro identifiant l'année	Clé étrangère
SURFACE_BATI	real	mesure de surface du bati	
TAUX	real	le taux de croissance du bati	
<i>RASTER</i>			
Champ	Type	Description	Remarque
RASTER ID	int	Numéro identifiant le raster	Clé primaire
SURFACE	real	la surface bati	
PERIMETRE	real	le perimetre de la zone bati	
NATURE	varchar	la nature du bati	
MAT_CONSTRCT	varchar	matériaux de construction	
ETAT	varchar	l'etat du bati	
GEOMETRY	varchar	les coordonnées géographiques	
<i>VECTEUR</i>			
Champ	Type	Description	Remarque
VECTEUR_ID	int	Numéro identifiant	Clé primaire
SURFACE	real	la surface bati	
PERIMETRE	real	le perimetre de la zone bati	
NATURE	varchar	la nature du bati	
MAT_CONSTRCT	varchar	matériaux de construction	
ETAT	varchar	l'etat du bati	
GEOMETRY	varchar	les coordonnées géographiques	
<i>TRASPORT</i>			
Champ	Type	Description	Remarque
TRANSPORT_ID	int	Numéro identifiant le transport	Clé primaire
TYPE_TRANS	text	le type de transport	

RES_ROUTIER			
Champ	Type	Description	Remarque
RES_ROUTIER_ID	int	Numéro identifiant du res-rte	Clé primaire
TRANSPORT_ID	int	Numéro identifiant le transport	Clé étrangère
STATISTIQUE_RTE	varchar	mesure de routes	
ANNEE_ID	int	Numéro identifiant l'année	Clé étrangère
COMMUNE_ID	int	Numéro identifiant la commune	Clé étrangère
TRAFIC	real	le taux de trafic routier	
TAUX	real	mesure reseau routier	
COMMUNE 1			
Champ	Type	Description	Remarque
COMMUNE_ID	int	Numéro identifiant la commune	Clé primaire
COMMUNE	text	le nom de la commune	
ZONE	text	la zone	
AGGLOMERATION	text	l'afflomération	
CITÉ	text	la cité	
STATISTIQUE_RTE			
Champ	Type	Description	Remarque
STATISTIQUE_RTE_ID	int	Numéro identifiant stat-rte	Clé primaire
STATISTIQUE	varchar	mesure statistique	
ANNEE			
Champ	Type	Description	Remarque
ANNEE_ID	int	Numéro identifiant l'année	Clé primaire
ANNEE	date	l'année	
MOIS			
Champ	Type	Description	Remarque
MOIS_ID	int	Numéro identifiant le mois	Clé primaire
MOIS	date	le mois	
JOUR			
Champ	Type	Description	Remarque
JOUR_ID	int	Numéro identifiant le jour	Clé primaire
JOUR	date	le jour	
DIM2			
Champ	Type	Description	Remarque
DIM2_ID	int	Numéro identifiant la dim 2	Clé primaire
ATT1	varchar	attribut a titre d'exemple	
CULTURE			
Champ	Type	Description	Remarque
CULTURE_ID	int	Numéro identifiant la culture	Clé primaire
CULTURE	text	attribut a titre d'exemple	
SITES	text	site culturel	

TOURISME			
Champ	Type	Description	Remarque
TOURISME_ID	int	Numéro identifiant le tourisme	Clé primaire
CULTURE_ID	int	Numéro identifiant la culture	Clé étrangère
STATISTIQUE TOUR_ID	int	Numéro identifiant stat-tourisme	Clé étrangère
ANNEE_ID	int	Numéro identifiant l'année	Clé étrangère
COMMUNE_ID	int	Numéro identifiant la commune	Clé étrangère
NBR_SITES	int	nombre de sites touristique	
HISTORIQUE	text	historique du sites touristique	
SOCIO_ECONOM			
Champ	Type	Description	Remarque
SOCIO ECONOM_ID	int	Numéro identifiant socio-econm	Clé primaire
DIM2_ID	int	Numéro identifiant dim2	Clé étrangère
DIM1_ID	int	Numéro identifiant dim1	Clé étrangère
ANNEE_ID	int	Numéro identifiant l'année	Clé étrangère
COMMUNE_ID	int	Numéro identifiant la commune	Clé étrangère
MESURE1	varchar	mesure a titre d'exemple	
MESURE2	varchar	mesure a titre d'exemple	
STATISTIQUE_TOUR			
Champ	Type	Description	Remarque
STATISTIQUE_TOUR_ID	int	Numéro identifiant stat-tourisme	Clé primaire
STATISTIQUE	varchar	mesure a titre d'exemple	
DIM1			
Champ	Type	Description	Remarque
DIM1_ID	int	Numéro identifiant la dim 1	Clé primaire
ATT1	varchar	attribut a titre d'exemple	

Bibliographie

- [AH87] SERGE ABITEBOUL and RICHARD HULL. IFO : A Formal Semantic Database Model. *ACM Transactions on Database Systems*, 12(24) :525–565, 1987.
- [ASCR] M. Adésir-Schilling, M-F. Courel, and S. Robert. le système d’information géographique.
- [Bé97] Yvan Bédard. Spatial OLAP. *2ème Forum annuel sur la RD, Géomatique VI : Un monde accessible*, 1997.
- [Bim07] Sandro Bimonte. *Intégration de l’information géographique dans les entrepôts de données et l’analyse en ligne : de la modélisation à la visualisation*. PhD thesis, L’Institut National des Sciences Appliquées de Lyon, 2007.
- [Bim08] Sandro Bimonte. L’information géographique et les entrepôts de données. *CFC N198*, 2008.
- [BLjPN04] Yvan Bédard, Suzie Larrivée, Marie josée Proulx, and Martin Nadeau. Modeling geospatial databases with plug-ins for visual languages : A pragmatic approach and the impacts of 16. In *Years of Research and Experimentations on Perceptory*, pages 17–30. Springer Berlin / Heidelberg, 2004.
- [BP89] Y. Bédard and F. Paquette. Extending entity/relationship formalism for spatial information systems. *AUTO-CARTO 9*, pages 818–827, 1989.
- [BPC92] Yvan Bédard, J Pageau, and C Caron. Spatial data modeling : The modular formalism and case technology. *ISPRS Symposium*, August 1-14 1992.
- [Bri02] Alex Brisebois. Notions avancées d’intégration de données dans un contexte de peuplement d’entrepôt de données spatiales. *SCG-65825 Sujet spécial*, 2002.
- [BTM06] S. Bimonte, A. Tchounikine, and M. Miquel. GeoCube, a Multidimensional Model and Navigation Operators Handling Complex Measures : Application in Spatial OLAP. *Proceedings of the 4th Biennial International Conference Advances in Information Systems*, 2006.

-
- [BTMR07] S. Bimonte, A. Tchounikine, M. Miquel, and R.Laurini. Vers l'intégration de l'analyse spatiale et multidimensionnelle. *SAGEO*, 2007.
- [BWTM06] S. Bimonte, P. Wehrle, A. Tchounikine, and M.Miquel. GeWOlap : A Web Based Spatial OLAP Proposal. *Springer-Verlag Berlin Heidelberg*, page 1596 – 1605, 2006.
- [Dan90] Tomlin C. Dana. Geographic information systems and cartographic modeling. page 572, 1990.
- [DS04] Jean Denègre and François Salgé. *Les systèmes d'information géographiques*, volume 2e édition. éditions PUF collection Que sais-je ?, 2004.
- [FI99] J. L. Filho and C. Iochpe. Um Estudo sobre Modelos Conceituais de Dados para Projeto de Bancos de Dados Geográficos. *Revista IP-Informática Pública*, 2(1) :37–90, 1999.
- [Hel92] C. Helen. People manipulate objects (but cultivate fields) : beyond the raster-vector debate in gis. *Springer Verlag*, pages 65–77, 1992.
- [Ie07] IT-expert. Le décisionnel. 68, juillet/août 2007.
- [Jé13] Laurent Jégou. Moteurs cartographiques et diffuseurs de webservices géospatiaux. novembre 2013.
- [KEL97] Sean KELLY. *Data warehousing in action*. John Wiley and Sons, 1997.
- [KL99] Juan-Luis Klein and Suzanne Laurin. *L'Éducation géographique : formation du citoyen et conscience territoriale*. 1999.
- [KRRW00] R. Kimball, L. Reeves, M. Ross, and W.Thornthwaite. Concevoir et déployer un data warehouse. *Editions Eyrolles*, 2000.
- [MBB02] Maryvonne Miquel, Yvan Bédard, and Alexandre Brisebois. Conception d'entrepôts de données géospatiales à partir de sources hétérogènes exemple d'application en foresterie. *Ingénierie des Systèmes d'Information*, 7(3) :89–111, 2002.
- [Mj01] Kraak Menno-jan. *Web Cartography*, volume 1. Francis and Taylor, New York, 2001.
- [MMB+96] Zakaria Maamar, Bernard Moulin, Yvan Bédard, D Vallière, and C Caron. Adapting data model for the design of spatio-temporal database. *Comp. Env. and Urban Systems*, 20(1) :19–41, 1996.
- [PdSB05] G. Pestana, M. Mira da Silva, and Y. Bédard. Spatial OLAP Modeling : An Overview Base on Spatial Objects Changing over Time. *IEEE 3rd International Conference on Computational Cybernetics*, 2005.

-
- [PMA01] R. Philippe, S. Michel, and V. Agnès. *AB initio* spatial databases with application to gis. *2nd Ed. San Francisco, CA*, page 440, 2001.
- [PMDD01] L. Paul, G. Michael, M. David, and R. David. Geographic information systems and science. page 517, 2001.
- [Ral96] KIMBALL Ralph. *The Data Warehouse Toolkit : Practical Techniques for Building Dimensional Data Warehouses*. New York : John Wiley and Sons, 1996.
- [SAA12] N. Selmourne, N. Abdat, and Z. Alimazighi. Towards a geographic data warehouse for water resources management. *IEEE Second International Workshop on Advanced Information Systems for Enterprises*, 2012.
- [SFP89] SFPT. Les sig. *Société Française de Photogrammétrie et de Télédétection*, 1989.
- [VAJ96] HARINARAYAN Venky, RAJARAMAN Anand, and ULLMAN Jeffrey. Implementing Data Cubes Efficiently. *ACM SIGMOD International Conference on Management of Data*, pages 205–216, 1996.
- [VBQ13] Alain Venot, Anita Burgun, and Catherine Quantin. *Informatique Médicale, e-Santé – Fondements et applications*. 2013.
- [VVSN12] Sangeeta Verma, Ravindra Kumar Verma, Anju Singh, and Neelima S. Naik. Web-Based GIS and Desktop Open Source GIS Software : An Emerging Innovative Approach for Water Resources Management . *Advances in Computer Science, Eng. Appl*, pages 1061–1074, 2012.
- [WD04] M. Worboys and Duckham. *Gis : A computing perspective*. *CRC, Boca Raton*, 2004.
- [W.H92] Inmon W.H. *Building the Data Warehouse*, volume 1. QED Information Sciences, Inc., Wellesley, 1992.
- [W.H96] Inmon W.H. *Building the Data Warehouse*, volume 2. New York : Wiley, 1996.
- [WIN98] Richard WINTER. Databases : back in the OLAP game. *Intelligent Enterprise Magazine*, page 60–64, 1998.