Environmental Health Systematic Review

How the new analytical geomatics technologies can help environmental health professionals and decision-makers to make further use of mapping than what is offered traditionally by geographic information systems (GIS) and web mapping.

For:
The National Collaborating Centre for Environmental Health (NCCEH)

Prepared by:
Marie-Josée Proulx, M.Sc.,
Eveline Bernier, M.Sc.,
Yvan Bédard, Ph.D
Centre de recherche en géomatique, Université Laval

People responsible:
Pierre Gosselin, MD, MPH
Institut National de Santé Publique du Québec

Yvan Bédard, Ph.D, full professor
NSERC Industrial Research Chair in Geospatial Databases for Decision Support

Review conducted as part of the project “Better Understanding Climate-Related Health Vulnerabilities” by the Canadian Network of centres of excellence in geomatics GEOIDE

November 30, 2007

Production of this report has been made possible through a financial contribution from the Public Health Agency of Canada through the National Collaborating Centre for Environmental Health. The views expressed herein do not necessarily represent the views of the Public Health Agency of Canada or the National Collaborating Centre for Environmental Health.

This document was originally written in French. It was translated in English by an external firm and the English version was reviewed by the team from Université Laval

We would like to thank the specialists in the health field who validated our document in order to adjust the language level to our readership.
Germain Lebel (Institut National de Santé Publique du Québec)
Alain Vanasse and Maria-Gabriela Orzanco
(Primus Group, Université de Sherbrooke)
Eddie Oldfield, Maurice Lanteigne and Xiaolun Yi
(New Brunswick Lung Association)

Appendix A on the various sets of spatial data of interest to environmental health has been enclosed.

Objectif du document :

Les organisations canadiennes investissent des centaines de millions de dollars annuellement dans l'acquisition de données sur le territoire, son occupation, ses ressources, ses habitants et son utilisation. D'un autre coté, les praticiens en santé environnementale ont développé une multitude de systèmes d'information dont plusieurs utilisent des systèmes d'information géographiques (SIG) et des technologies de cartographie sur le web pour supporter leurs différentes activités. L'utilisation des SIG en santé publique a mené l'organisation de plusieurs conférences ces 15 dernières années aussi bien que la rédaction de publications scientifiques. Plusieurs projets ont atteint un haut niveau de maturité. Cependant, ces systèmes ont été développés sur des technologies géospatiales qui sont 'transactionnelles' de nature et qui ne bénéficient pas des avancées offertes par les plus récentes technologies décisionnelles propres au domaine appelé « Business Intelligence » (BI).

Par leur nature, les technologies BI ne sont pas construites pour gérer les transactions; elles sont construites pour supporter les analyses complexes et la découverte de connaissances. Le BI est basé sur une structure de données différente, appelée hypercube, et inclut des technologies comme les tableaux de bord, l'OLAP (on-line analytical processing), le forage de données, les marchés de données et les entrepôts de données. De nos jours, l'usage des technologies BI est devenu une pratique courante dans plusieurs organisations puisque ces technologies sont disponibles commercialement depuis une décennie. Cependant, ce n'est que depuis quelques années à peine que les logiciels commerciaux permettent le couplage des technologies BI aux données géospatiales. Même si la recherche universitaire est active dans ce domaine depuis le milieu des années 90, ce n'est que récemment que des solutions intégrées sont disponibles sur le marché par des petites compagnies innovatrices aussi bien que des fournisseurs de logiciels majeurs.

Ces nouvelles technologies permettent d'une manière beaucoup plus efficace (rapidement et facilement) ce qui est extrêmement difficile et long avec les technologies SIG et de cartographie sur le web, comme produire des informations agrégées, produire des données de synthèse, des analyses de tendance, des comparaisons spatio-temporelles, de l'analyse interactive, de la découverte de connaissances, etc. Ces technologies ne sont pas dédiées à remplacer les SIG et les technologies de cartographie sur le web, mais plutôt à permettre de nouvelles analyses sur les données disponibles dans les systèmes transactionnels déjà en place et par conséquent offrir un meilleur retour sur l'investissement.

Depuis que les organisations en santé publique acquièrent un volume significatif de données complexes et ont besoin de systèmes pour consulter et analyser les tendances reliées aux expositions environnementales et aux problèmes de santé, elles doivent fournir à leurs spécialistes en santé les technologies géospatiales décisionnelles les plus efficaces. De plus, cela doit leur permettre d'accéder aux informations appropriées (cf. meilleur usage) d'une meilleure manière (cf. plus rapidement). Cela est vrai pour les besoins en planification en santé publique, la gestion et la surveillance en général.

De façon à aider les praticiens et politiciens en santé environnementale à évaluer les possibilités et difficultés associées aux technologies BI géospatiales émergentes, nous proposons ce rapport. Dans ce rapport, nous nous concentrerons sur les technologies suivantes parmi l'ensemble des technologies géodécisionnelles: tableaux de bord spatiaux, OLAP spatial, forage de données spatiales et les entrepôts de données spatiales. Conséquemment, ce rapport n'inclut pas une revue systématique des technologies SIG traditionnelles des praticiens en santé.

Document objective:

Canadian organizations invest hundreds of millions of dollars annually to acquire data about the land and its occupation, resources, inhabitants and uses. In the same way, practitioners in environmental health have developed several information systems, most of them using GIS (Geographical Information Systems) and Web-mapping technology to support their different activities. Combining public health and GIS has led to the organization of several symposia over the last 15 years as well as scientific publications. Many projects have reached a high level of maturity. However, these systems have been developed with geospatial technologies that have a "transactional" nature and they don't benefit from the advances offered by the most recent decision-support technologies offered in the field called "Business Intelligence" (BI).

BI technologies aren't built to manage data transactions; they are built to support complex analysis and knowledge discovery. BI relies mostly on a different data structure, called hypercube, and encompasses technologies such as Dashboards, OLAP (On-Line Analytical Processing), Data Mining, Datamarts and Data Warehousing. Nowadays, using BI technology has become common practice in several organizations since these technologies have been commercially available for over a decade. However, it is only within a few years that commercial software has appeared to allow users to bridge BI and geospatial technologies. In spite of university R&D going back to the mid-90s, it is only recently that integrated solutions have been made available on the market by small innovative companies as well as major software providers.

These new technologies allow one to perform in a more efficient manner (faster, easier) what is very difficult and time-consuming with typical GIS and web-mapping technologies, namely to produce summarized information, aggregated data, trends analysis, spatio-temporal comparisons, interactive exploration of data, geographic knowledge discovery, etc. Such technologies are not meant to replace GIS and web-mapping applications, but rather to add new analysis capabilities on data stored in actual systems and provide a better ROI (return on investment).

Since public health organizations collect significant volumes of complex data and need systems to monitor and assess the trends related to environmental exposures and related health problems, they need to provide their health specialists with the most efficient geospatial decision-support technologies. Furthermore, this may allow more of them to access appropriate information (cf. better ease-of-use) in a timelier manner (cf. much faster response times). This is true for public health planning, management and surveillance purposes in general. Some basic organizational requirements also need to be met for a successful implementation of those technologies.

In order to help environmental health practitioners and policymakers to assess the possibilities and difficulties associated with the emerging geospatial BI technologies, we propose this report. In this report we focus on the following technologies: spatial dashboard, spatial on-line analytical processing (SOLAP), spatial data mining, and spatial data warehouse. Consequently, the report does not include a systematic review of traditional GIS of health practitioners.

TABLE OF CONTENTS

List of figures..9
List of tables 11
Acronyms in the health field 12
Acronyms in geomatics and informatics 14
1. General Introduction.. 16
1.1 The Power of Maps.. 16
1.2 Geomatics Technologies 18
2. Methodology 20
3. Background..22
3.1 The Transactional Systems 22
3.2 Decision-support Systems..24
4. The Transactional Systems 26
4.1 Geographic Information Systems..26
4. 1. a. Summary Definition 26
4. 1. b. State of the Art 27
4. 1. c. Existing Technologies 29
4. 1. d. Examples of Health Applications 29
4.2 Web Mapping... 31
4. 2. a. Summary Definition 31
4. 2. b. State of the Art 31
4. 2. c. Positioning in Relation with Other Technologies 33
4. 2. d. Existing Technologies 34
4. 2. e. Examples of Health Applications 34
i. Examples of Web Mapping Applications 36
ii. Examples of GIS Applications on the Web...37
iii. Examples of Digital Atlas Applications 39
5. Analytical systems 44
5.1 spatial data warehouses...46
5. 1. a. Summary Definitions.. 46
5. 1. b. State of the Art 46
i. Concepts of the Non-Spatial World... 46

ii. The Concepts of the Spatial World ... 49
5. 1. c. Existing Technologies ... 50
5. 1. d. Examples of Health Applications ... 51

5.2 spatial on-line analytical processing (OLAP) ... 53

5. 2. a. Summary Definitions ... 53
5. 2. b. State of the Art ... 53
i. Concepts of the Non-Spatial World ... 53
ii. Concepts of the Spatial World ... 57
5. 2. c. Positioning in Relation to Other Technologies ... 58
5. 2. d. Existing Technologies ... 60
i. OLAP Technologies: ... 60
ii. SOLAP Technologies: ... 61
5. 2. e. Examples of Health Applications ... 65

5.3 spatial dashboard ... 68

5. 3. a. Summary Definitions ... 68
5. 3. b. State of the Art ... 68
5. 3. c. Positioning in Relation to Other Technologies ... 69
5. 3. d. Existing Technologies ... 72
5. 3. e. Examples of Applications in Health ... 73

5.4 spatial data mining ... 76

5. 4. a. Summary Definitions ... 76
5. 4. b. State of the Art ... 76
5. 4. c. Spatial Data Mining ... 78
5. 4. d. Positioning in Relation to Other Technologies ... 79
5. 4. e. Existing Technologies ... 80

6. future research and development perspectives ... 82
7. Conclusion ... 85
8. References ... 87

LIST OF FIGURES

Figure 1. Positioning of operational systems in the deployment of a corporate architecture for decision-making 26

Figure 3. Example of corvidae reporting presented in the Système intégré des données de vigie sanitaire du virus du Nil occidental (SIDVS-VNO). 39

Figure 4. Interactive SIRSé map presenting the comparative rates of breast cancer mortalities in women by nearby territories, for the 1995-1999 periods. 40

Figure 5. Pennsylvania Cancer Atlas presenting the incidence of colorectal cancer by county, for all races, for the 1994-2002 periods. 41

Figure 6. Breast cancer mortality rate map for Quebec along with the table of associated data 42

Figure 7. Map of breast cancer mortality rates in Canada along with the table of associated data 43

Figure 8. Illustration of a multidimensional cube representing the incidence rate based on territory, time and diseases. 45

Figure 9. Positioning of transactional systems, data warehouses and data marts (Bédard *et al.*, 1997). 47

Figure 10. Data Warehouse Architectures 48

Figure 11. Positioning of data warehouse architecture examples in the deployment of a corporate architecture dedicated to decision-making 49

Figure 12. The Three Types of Spatial Dimensions (source: Rivest *et al.*, 2003) 50

Figure 13. Illustration of the various OLAP server architectures (source: Bédard et al., 1997) 56

Figure 14. Positioning of the tools based on the nature of the information and its aggregation levels (source: Proulx et Bédard, 2004b). 58

Figure 15. Positioning of the GIS capacities and a SOLAP tool (source: Proulx et al., 2004a). 59

Figure 16. Positioning of possible SOLAP technology couplings in the deployment of a corporate architecture dedicated to decision-making. 65

Figure 17. Positioning of the spatial dashboard in the deployment of a corporate architecture dedicated to decision-making. 72

Figure 18. Strategic dashboard by the Carrefour montréalais d'information sociosanitaire presenting the requests for accommodations by people 75 and older in three CSSS. 74

Figure 19. Application for map production management developed by the Université Laval and Syntell for the MRNFQ. 74

Figure 20. Data mining is part of the global knowledge discovery in database (KDD) (Han & Kamber, 2001) 77

Figure 21. The principal fields that are the basis of spatial data mining 79

Figure 22. Positioning of spatial data mining in the deployment of a corporate architecture dedicated to decision-making. 80

LIST OF TABLES

Table 1. Comparison between the operational systems, data warehouses and data marts (Bédard *et al.*, 1997). 47

Table 2. Comparison between the characteristics of relational and multidimensional OLAP architectures (source: Proulx *et al.*, 2004). 56

Table 3. Principal distinctions between the use of a GIS and an OLAP in regards to the users targeted, the type of data, the processes and architectures supported (source: Proulx & Bédard, 2004b). 60

Table 4. OLAP Servers and their Client Tools 61

Table 5. Positioning of dashboards in relation to OLAP tools 71

Table 6. Evolution in the field of databases (Thearling, 2007). 77

Table 7. Comparison of the spatial analysis and spatial data mining (Zeitouni, 2006).......79

ACRONYMS IN THE HEALTH FIELD

AQI	Air Quality Index
Catch	Comprehensive Assessment for Tracking Community Health
CDC	Center for Disease Control and Prevention
CERN	European Organization for Nuclear Research
CIHI	Canadian Institute for Health Information
CSSS	Centre de santé et de services sociaux
INSPQ	Institut National de Santé Publique du Québec
LIBCSP	Long Island Breast Cancer Study Project
MeSH	Medical Subject Headings
NBLA	New Brunswick Lung Association
NCCEH	National Collaborating Centre in Environmental Health
NCD	Non-communicable diseases
NCI	National Cancer Institute
NHSI	National Health Surveillance Infrastructure
NIEHS	National Institute of Environmental Health Sciences
NIH	National Institutes of Health
NLM	National Library of Medecine
PHAC	Public Health Agency of Canada
PIRISP	Plan intégré des ressources informationnelles en santé publique [Integrated Plan for Public Health Informational Resources]
SIDVS	Système intégré des données de vigie sanitaire [Integrated System for Health Surveillance Data]
SIRSé	Système d'information régional en santé [Regional Health Information System]
SIST-MC	Système d'Information Spatio-Temporel sur les Maladies Chroniques
WHO	World Health Organization
WNV	West Nile Virus

ACRONYMS IN GEOMATICS AND INFORMATICS

AJAX	Asynchronous JavaScript
API	Application Programming Interface
BI	Business Intelligence
CSIRO	Commonwealth Scientific and Industrial Research Organisation
DBMS	Database Management Systems
EIS	Executive Information System
ESRI	Environmental Systems Research Institute
ETL	Extraction, Transformation and Loading
GIF	Graphics Interchange Format
GIS	Geographic Infomation System
GMQL	GeoMiner Query Language
GOLAPA	Geographic On-line Analytic Processing Architecture
GPS	Global Positioning System
HOLAP	Hybrid OLAP
HTML	Hypertext Markup Language
ICDG(CGDI)	Canadian Geospatial Data Infrastructure
ICEM/SE	Interface Cartographique pour l'Exploration Multidimensionnelle des indicateurs de Santé Environnementale sur le World Wide Web [Mapping Interface for the Multidimensional Exploration of Environmental Health Indicators on the World Wide Web]
IT	Information Technology
JDBC	Java Database Connectivity
JPEG	Joint Photographic Experts Group
KDD	Knowledge Discovery in Databases
LBS	Location-Based Services
LIDAR	Light Detection and Ranging
MDX	Multidimensional Expression
MOLAP	Multidimensionnal OLAP

ODBC	Open Database Connectivity
OGC	Open Geospatial Consortium
OLAM	On-line Analytical Mining
OLAP	On-line Analytical Processing
OLTP	On-line Transactional Processing
OMG	Common Warehouse Metamodel
PNG	Portable Network Graphics
RDBMS	Relational Database Management System
ROLAP	Relational OLAP
SOLAP	Spatial On-line Analytical Processing
SOVAT	Spatial OLAP Visualization and Analysis Tool
SQL	Structured Query Language
SVG	Scalable Vector Graphics
TDWI	The Data Warehousing Institute
UMapIT	Unrestricted Mapping Interactive Tool
USGS	United States Geological Survey
W3C	World Wide Web Consortium
WFS	Web Feature Service
WMS	Web Map Service
WebGIS	Geographic Information System on the Web
XML	Extensible Markup Language

1. GENERAL INTRODUCTION

1.1 THE POWER OF MAPS

With the recent developments in geomatics sciences (high-resolution satellite imaging for monitoring the environment, precise dynamic positioning in real-time using GPS satellites for emergency systems, LIDAR airborne laser system for measuring the topography and its movements, geographic information systems on the web to map and analyze socio-economic data on the population, etc.), we now produce terabytes of geo-referenced data at a much more affordable cost than just a decade ago. One need only think of Google Map, MapQuest and Virtual Earth (Microsoft) to notice it the moment you go on the Internet. Such technologies have become omnipresent and have now made their way into the health field: epidemiology, kinesiology, home services, ambulance services, etc. There is even a health geography community with its own symposiums and publications.

Most of the health data already has a spatial component such as addresses, location names, postal codes, connection to the local health centres (e.g. the Community Health Centre Areas in Quebec), or in certain cases geographic coordinates (latitude and longitude). The same is true for environmental data and population data, which is often mapped (e.g. Statistics Canada data). As a general rule, there are rather simple applications that use map data from a single source, often not very up-to-date, sometimes incomplete and providing limited precision. Such maps can be good enough for several usages. However, the needs can be more complex, both in terms of the quality and quantity of the maps required as well as the diversity of the data sources, the coverage of various time periods and the taking into account of various levels of analysis (local, regional, provincial, national and international) for comparative, summarizing and forecasting purposes (typical of decision-making processes).

Spatial data provides the position, the shape and distribution of phenomena on our territories, whether these are administrative, political or natural. These localized phenomena number in the thousands and make it possible to better understand the spatial correlations between the environment and people's health, the evolution in drinking water quality in relation with the evolution in the use of the land, the distribution of emergency vehicles in real-time or the economic data on families by neighbourhood for example. Viewing phenomena on a map makes it easier to see their spatial properties and facilitates comprehension. This includes the intrinsic spatial characteristics of the phenomenon (position, form, size, orientation, direction of movement, etc.), spatial relationships (adjacency, connectivity, inclusion, proximity, exclusion, superposition, etc.) and spatial distribution (concentrated, by groupings, regular, etc.). For example, in viewing a map representing various regions, it allows us to make comparisons. By viewing various maps from the same region, it allows us to discover correlations between the phenomena. And lastly, by viewing a region based on various time periods, it allows us to assess the phenomena's evolution. When we use a map, we end up with a more complete understanding of a phenomenon's structure and its relationships with other phenomena than what we can get when using a table or diagram. When we combine maps with statistical tables and diagrams, it can all be compared to discover new knowledge.

Maps naturally help in the process of discovering knowledge. Within a context of spatial data, maps do more than make data visible; they become active tools in supporting the user's thought process. Maps can display information that would not be visible using table data, especially for phenomena whose spatial distribution does not follow pre-established boundaries (e.g. administrative boundaries). Several studies in the cognitive sciences have demonstrated the superiority of images over numbers and words in stimulating understanding and memory (Buzan & Buzan, 2003; Fortin & Rousseau, 1989; Standing, 1973), resulting in a better knowledge discovery process (brain more alert, better visual rhythm and better general perception).

However, the full potential of maps remains untapped because the technologies traditionally used to produce maps are not very good at supporting the cognitive processes of decision-makers and analysts who want to understand a phenomenon, issue hypotheses and discover new knowledge. For example, as geographical information systems do not intrinsically support temporal analysis operators (although certain universal servers provide some such operators), the temporal tendency analysis remains a complex and time-consuming function to perform in such systems. It is largely recognized in the literature that these characteristics (complex and time-consuming) make traditional GIS tools less adapted for the decision-making process (see for example Turban *et al.*, 2005).

Until recently, few major solutions have been proposed that would go beyond the traditional solutions of GIS and web mapping. Today, the coupling of geographic information systems (e.g. ArcGIS) or universal servers (e.g. Oracle Spatial) with business intelligence technologies (e.g. OLAP servers and dashboards in particular) makes it possible to make better use of cartography and to go beyond the traditional technologies (in a complementary manner, without looking to replace them). GIS remain the primary tools for performing spatial analyses (e.g. spatial correlations, optimal routes, density analysis) while the new spatial analytical systems facilitate the consultation and analysis of the results and allow to compare them and derive new analyses.

1.2 GEOMATICS TECHNOLOGIES

There are dozens of health applications deployed on geographic information systems (GIS) in Canada and there are even more around the world. In 2001, a study (Boulos *et al.*, 2001) had counted twenty or so GIS applications in health. In 2004, the author conducted a new review and updated with a long list of GIS health applications (Boulos, 2004).

However, the vast majority of developed applications come from the world of transactional systems (which includes GIS) and few professionals have expanded their applications beyond these systems. The report by the New Brunswick Lung Association (2006) on Canadian geo-spatial infrastructures states: "*There is no other technology that can merge the necessary data sets into a meaningful representation (exposing relationships, dependencies, etc.) than web-based GIS.*" This statement was not based on a study of the possibilities offered by the new spatial analytical technologies and rather limited itself to transactional technologies, which is representative of the current market situation. Like most software users, health professionals have a natural tendency to rely upon the tools with which they are familiar. Moreover, most GIS stakeholders have a lack of knowledge regarding analytical technologies. Consequently, they often attempt to employ the transactional technologies that they are accustomed to, i.e. GIS, to meet their decision-making needs, despite the fact that the geomatics community is increasingly acknowledging that these tools cannot adequately satisfy their needs (Bédard *et al.*, 2003). However, in a manner that is very representative of the current situation, the New Brunswick Lung Association has since that time invested in order to upgrade their GIS application towards their spatial analytical needs through internal IT development, often using open source solutions and making use of international standards (e.g. OGC, W3C). Therefore, across the country and abroad, several teams have had to do some important development work and invest considerable amounts to partially complete the capacities of their geographic information systems to support decision-making.

As Viatis and Tzagarakis (2005) stated so well: "*These systems are transaction-oriented and like every transactional system (e.g. database management systems), they do not efficiently address summarized information, cross-referenced information and interactive exploration of data. Furthermore, GIS are efficient to deal neither with temporal data nor with multiple levels of data granularity.*"

Consequently, several digital atlas applications have been developed (see section 4.2.e.iii) in this community in order to respond to decision needs. However, these applications are rarely deployed with analytical geomatics technologies, because they have only just recently appeared on the market and are not well known. Rather, these applications are being developed with classic GIS and web mapping tools. It would be better to use them with more appropriate and high-performance technologies based on the multidimensional paradigm (i.e. data cubes paradigm). In fact, an array of tools from the business intelligence world can meet the new needs of health professionals when used with geomatics technologies. Curiously, these technologies are often in place in health organizations, but in sectors of activity other than those where GIS are deployed. Business intelligence and geomatics have therefore evolved at the same time, without ever coming face to face, both in health organizations and in other organizations. It has only been recently that, after ten years of research in universities and experiments in organizations, the first "analytical geomatics" commercial solutions saw the light of day.

This literature review is presented as follows. First, the methodology used for this report is detailed in the next section, followed by the presentation of the two major categories of systems that are the subject of this study (Chapter 3). A picture of how transactional systems, such as GIS and web mapping, are used in the health field is presented in the first section and analytical geomatics solutions are summarized in the second. A detailed presentation of the transactional and spatial BI solutions is the subject of chapters 4 and 5 respectively. The sixth chapter focuses on the future outlooks for spatial BI technologies whereas the conclusion is the subject of chapter 7. All of the bibliographical references are in chapter 8 and an appendix has been enclosed to present examples of mapping data sources that are potentially interesting for the development of environmental health applications.

2. METHODOLOGY

The methodology used to prepare this document first consisted in a review of the scientific publications pertaining to the field of health and geographic information systems, which took place in January and February 2007. Document search sites such as CiteSeer Publications Research Index[1] and Current Contents[2] were queried using basic keywords (GIS, Geography, Geographic Information Systems and Health), making it possible to identify publications combining GIS and health. Several publications presented in the Annual Review of Public Health and the International Journal of Health Geographics dealing with geographic information systems served as a point of departure to position GIS in the health world. Several of their references on existing applications were then presented and updated.

After reading these first publications of interest, research was conducted through the networking of references listed in order to complete the information on the various applications used by the health organizations presented in these publications (World Health Organization, National Cancer Institute, USGS National Atlas). A keyword search on health terms (e.g. Epidemiology, Public Health, Health Care, Mortality, Cancer, Emergency Response, Disasters, Disease and Disease Surveillance) completed the list of applications.

Then, a search of the *GeoConnections*[3] site made it possible to identify several application projects in the health world (e.g. Public Health Agency of Canada, World Health Organization, the New Brunswick Lung Association) financed by this organization and promoting the deployment of geomatics applications and the use of standards.

Several reports and articles published by the senior author were used to build sections 4 and 5 on the concepts pertaining to the transactional and decision-support fields. In fact, the Research Chair, the holder of which is the senior author, conducted a review of the GIS, DBMS, OLAP, dashboard, SOLAP and the spatial dashboard on a continual basis for years. A few years ago, this review was the basis of a needs study conducted for Health Canada (see Mowat *et al.*, 2000). However, a new Internet search made it possible to improve the list of existing applications in the field of health using a list of keywords from the decision-support field (combining spatial with data warehouse, OLAP, on-line analytical processing, geovisualization, data mining, and dashboard).

[1] *http://citeseer.ist.psu.edu/*
[2] *http://www4.bibl.ulaval.ca/cc/*
[3] *http://www.geoconnections.org*

A primary draft of this literature review, written in French, was submitted to the NCCEH on March 31. This version of the document was then sent to three Canadian groups using and developing health applications for review and comments. The reviewers validated the document's language level, identified anything that was forgotten or if necessary any inappropriate tangents. The selection of the reviewer groups was done based on the study of the applications that we inventoried in order to select three different reviewer profiles. First, the Institut National de Santé Publique du Québec was selected for its leading edge status over the last several years in the deployment of decision-support geomatics applications through its participation with GEOIDE, the Canadian Network of centres of excellence in geomatics. The PRIMUS group at the Université de Sherbrooke is a public health research group that is being innovative in the deployment of decision-support applications since their recent involvement in GEOIDE. Lastly, the New Brunswick Lung Association, very active with *GeoConnections* and which largely adopts international standards, became a good reference for validating the current use of GIS in the health sector as well as their limits in decision-making.

Consequently, interviews were conducted with the following people to gather their comments:

- On May 31, 2007, with Mr. Germain Lebel of the Institut National de Santé Publique du Québec.
- On June 1, 2007, with Ms. Maria-Gabriela Orzanco and on June 26, 2007, with Mr. Alain Vanasse of the PRIMUS Group at the Université de Sherbrooke in Quebec.
- On June 20, 2007, with Mr. Eddie Oldfield, Mr. Maurice Lanteigne and Mr. Xiaolun Yi of the New Brunswick Lung Association.

The comments from these reviewers were discussed and integrated in the first version of the document.

Finally, Health Canada reviewers were identified and have expressed comments on the document, which were introduced in the second version of this report produced in November 2007.

3. BACKGROUND

3.1 THE TRANSACTIONAL SYSTEMS

The **transactional systems,** which date back to the early 70s, are designed so as to efficiently manage transactions and their impacts on the underlying databases. In the literature, these systems are often referred to as *On-line Transactional Processing (OLTP)*. They are used to manage all kinds of transactions, whether they are of a financial nature (e.g. banking systems), a commercial nature (e.g. inventory systems), a tourism nature (e.g. centralized reservation systems) or others. This type of system places emphasis on the storage of, access to and updating of data and its integrity management. Strategies must also be implemented in order to manage various technical aspects such as the processing of concurrent transactions conducted by several users simultaneously. The transactional systems rely on databases, the contents of which are typically very detailed and current. The updating of a transactional system often results in the replacement of given data by new data, the old data being archived rather than being kept online in the system. Queries into the history of a situation therefore often become difficult. Most transactional systems rely on a database implemented in accordance with a standardized relational structure that ensures minimum redundancy and help in maintaining the coherence of the data. Generally, an organization has several transactional systems and each of them is designed to manage a specific aspect of the organization (e.g. human resources, sales, production). Sometimes, coherence is assured through the implementation of an ERP solution (Enterprise Resource Planning) in the form of SAP or PeopleSoft. The principal tools in the transactional world are Database Management Systems (DBMS) such as Oracle, SQL-Server and PostgreSQL and then for geographic data, GIS such as ArcGIS, MapInfo and PostGIS or online mapping software such as MapServer, JMap or Geomedia WebMap.

Although they are efficient in the processing of transactions, transactional systems are not optimized for data analysis in a decision-making context. First of all, their standardized relational structure complicates and slows down the process of complex queries. In fact, these systems are optimized to support a large number of queries, each time involving a small quantity of data (e.g. number of hospitalizations for a person). However, a business-intelligence type query generally calls on a large number of parameters and must search a significant portion of the database to process a large number of data (e.g. number of people between the ages of 40 and 65, in the province of Quebec, who have cancer, for the years 2000 to 2005). This type of query typically requires a large number of joins between the various tables in the database, resulting in complicating the definition of queries and to reduce the performances. Secondly, business-intelligence type analyses generally require summary, aggregate or summarized data (e.g. cancer rates in a region for a given period). However, the transactional systems are mostly composed of very detailed data (e.g. the information on each cancer case) whereas several business-intelligence applications are composed mostly of aggregated and pre-calculated data (resulting in accrued performance). Thirdly, due to their dispersion within an organization and their independence, it is difficult, even impossible, to obtain a general picture of the situation being studied by the organization. In order to do so, these systems must be connected in an *ad hoc* manner, in other words they must be integrated in a certain way so they can be queried. Fourthly, strategic decision-making regularly requires temporal analyses in order to determine a phenomenon's evolution through time and therefore requires the presence of both current data and historic data, which the GIS have a hard time supporting. Lastly, in order to adequately support strategic decision-making, the system must offer analytical capabilities that are easy and fast in order to deal with unpredictable queries, but typically using the same analysis themes. Unfortunately, the GIS, due to their complexity (querying language, the need to know the internal structure of the database, specialized interface) do not offer the user-friendliness sought or the desired performance for this type of query. In short, the principal limitations of the transactional systems in efficiently supporting decision-making are:

- Their optimized relational structure for managing data, short transactions and ensuring the integrity of data rather than for facilitating their analysis and supporting complex queries;
- Their content composed of detailed and current data (often without a history that is directly accessible) rather than aggregated and summarized data supporting comparisons over time;
- Their multitude, dispersion and independence within companies;
- Their querying complexity (querying language and specialized interface).

In the mid 90s, from these problems came a new category of systems, referred to as decision-support or analytical. Contrary to the transactional systems, they are basically designed to support decision-making at a strategic level. However, it is important to specify that **decision-support systems** do not aim to replace the transactional systems already in place within the organizations. Rather they are complementary by allowing a completely other type of operation and analysis. They are generally designed so as to use the data from transactional systems already in place, to bring in new capacities that are not available in existing systems (typically transactional). Usually, analytical systems are developed within organizations that already possess transactional systems; however, it is possible to develop decision-support systems directly from data that is external to the organization, without having to first implement transactional systems.

3.2 DECISION-SUPPORT SYSTEMS

In order to optimize analyses, **decision-support systems** generally rely on a completely different paradigm than the one used for transactional systems, i.e. they rely on the multidimensional paradigm (or data cube). The structures based on this paradigm have demonstrated, over the last thirty years, that they were easier for users to understand than the transactional structures and therefore easier to query (Kimball, 2002a). Specialized tools exist to effectively explore and analyze data stored in such structures. These tools can be categorized in three principal categories: requesters and dashboards, *on-line analytical processing (OLAP)* tools and *data mining* tools.

There is a real need, in the health surveillance field, for business-intelligence applications, according to Cromely (2003): *"There is a need for disease surveillance systems that are focused on distributing information for meaningful spatial aggregates that meets the needs of the larger research community and the general public."* Consequently, these new technologies could favour the deployment of applications in the health field that are more useful than those based on traditional web mapping and GIS.

Even if the combination of digital and spatial data in the health research sector offers a large potential for providing experts with the capacity to discover new knowledge, the business-intelligence systems are extremely rare in this field (Mowat *et al.*, 2000), even today. A report was prepared in 2006 by the *New Brunswick Lung Association* (NBLA, 2006) in order to assist *GeoConnections* and the Canadian public health authorities in improving the mapping technologies for decision-making. The Association stated that since this wish was expressed, only three research groups in Canada followed the recommendations of the initial proposal by initiating health projects, i.e. at the University of Toronto (Buckeridge *et al.*, 2002) and McMaster University (Elliott *et al.*, 2001), which developed GIS applications, and a team from the Université Laval (Bédard *et al.*, 2003), which developed a spatial OLAP application.

Furthermore, other innovative applications, following the business-intelligence approach, saw the light of day over the last few years elsewhere in the world through university collaborations (in addition to the Canadian project at the Université Laval (Bédard *et al.*, 2003)). Two projects presenting the use of the OLAP approach for decision-making, i.e. the University of Ljubljana in Slovenia (Hristovski *et al.*, 2000) and another example by the University of Pittsburgh (Scotch & Parmanto, 2005) goes further by allowing the coupling between OLAP and spatial data.

It is therefore based on this new reality, which reflects the beginning of the use of business-intelligence systems in health that this literature review will successively present the positioning of the current applications in the health field as well as innovation solutions for the medical scientific community. It is then with the vision of a new approach that this report will illustrate, through concepts and examples, that health professionals should go beyond classic GIS and web mapping in order to improve decision-making, i.e. by planning the deployment of analytical applications based on new technologies and approaches such as spatial OLAP and spatial dashboards.

In the next chapters, we will successively present a literature review on *geographic information systems* (section 4.1), *web mapping* (section 4.2) and *data warehouses* (section 5.1). We will then address the *analytical concepts*, laying the necessary groundwork for the presentation of the analytical tools that are of interest to environmental health in the subsequent sections, i.e. *Spatial OLAP* tools (section 5.2), *spatial dashboards* (section 5.3) and *spatial data mining* (section 5.4). Each decision-support tool will be presented based on its spatial and non-spatial variant. For each tool, we will present definitions, its state of the art and existing technologies. Each tool will be positioned in relations with the others in order to help the reader in understanding the characteristics unique to each. Lastly, the health applications relative to these various tools, which were identified by our researchers using the Internet and in our communications with Canadian health groups, will be presented. Unfortunately, since not all developments have been published or disseminated, it is quite probably that certain Canadian projects and applications have not been included in this literature review. However, we believe we have covered a sufficient number of applications in order to illustrate and support our arguments and to ensure that the reader is capable of assimilating the major characteristics of the technologies presented. Lastly, we will discuss the research perspectives unique to these tools in geospatial decision-support systems (section 6).

4. The Transactional Systems

Several current systems within organizations are part of the transactional systems category: database management systems (DBMS), universal servers (DBMS supporting the spatial component of data) and geographic information systems (GIS) are popular examples. When they are used as source systems in the deployment of a corporate architecture, possibly supporting decision-making tools, they are called operational systems. The following figure positions these examples of operational systems in the deployment of a corporate solution dedicated to decision-making.

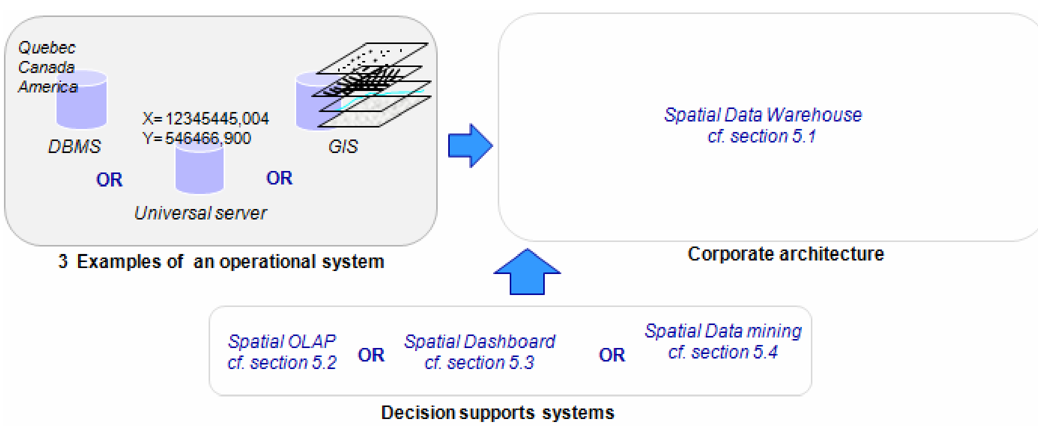


Figure 1. Positioning of operational systems in the deployment of a corporate architecture for decision-making

4.1 Geographic Information Systems

4. 1. a. ***Summary Definition***

A ***Geographic Information System*** can be defined as "a computer system capable of capturing, storing, analyzing, and displaying geographically referenced information; that is, data identified according to location. Practitioners also define a GIS as including the procedures, operating personnel, and spatial data that go into the system." (United States Geological Survey, USGS)[4]

In English, the acronym ***"GIS"*** is used freely. It is used to designate the organizational information system that processes the geographic data, the software included in the system and sometimes, and wrongly, the discipline of geomatics (see GIScience).

[4] *http://erg.usgs.gov/isb/pubs/gis_poster/#what*

4. 1. b. ***State of the Art***

In Ottawa, Ontario, at the Department of Energy, Mines and Resources, the first geographic information system was created. Developed by Roger Tomlinson in 1964, it was named the *"Canadian Geographic Information System"* and was used to store, analyze and manipulate data generated by the *Canada Land Inventory*. Afterwards, growth in the 1990s stimulated the use of GIS on Unix workstations and personal computers.

Curiously, even though this technology dates back to the 1960s and became a commercial success in the mid-80s, it was only in 2003 that the bilingual thesaurus (MeSH) published by the National Library of Medicine (NLM) included the term Geographic Information System (GIS)[5]. Although several administrative constraints can delay a term's entry into the thesaurus, its importance and usage in the health field has only been very recent.

Historically, health data was distributed in the form of printed atlases where the authors carefully selected the maps to be presented. Gradually, printed atlases gave way to digital and downloadable atlases over the Internet (e.g. in Acrobat format). Recently, an increase in the use of GIS and the Internet has resulted in several atlases now being available in the form of cartographic interfaces on the web, enabling users to produce dynamic maps that are of interest to them using available data and themes (Bell *et al.*, 2006).

GIS quickly become of interest to health managers thanks to their ability to display and conduct spatial analyses. Spatial data is of interest to managers because many of their decisions are based on geographic information and administrative constraints. Despite the significant advances in GIS over the last ten years, the observation made by De Lepper *et al.* (1995) still remains valid today, i.e. that the algorithms developed for the geomatics community does not systematically meet professional health needs since these systems are driven by the world of science and technology instead of the social sciences. Adaptations may be necessary and thereby delay the adoption of the technologies in question. However, health professionals are generally not exposed much to GIS technologies and this is mainly due to a misunderstanding of the technologies, their access and financial constraints (Jaishankar & Jhonson, 2006). Therefore, several writers in the public health field point out the underutilization of GIS functions in health (Boulos, 2004; Mowat *et al.,* 2000; Freier, 2000). Mowat pointed out, that for Canada, when GIS are used in the public health field, they are mainly used for basic functions such map displaying. *"In relation to the use of GIS, a survey of 30 community health researchers indicated that 70% were aware of GIS and felt that it could aid them in decision making, however most felt that it*

---

[5] *http://ist.inserm.fr/basismesh/mesh.html*

was not being utilized to its full potential, using it mostly for just presentation and visualization of spatial data." Mowat *et al.* (2000)

"Disease mapping is one of the branches of geographical epidemiology fulfilling the need to create accurate maps of disease morbidity and mortality. For instance, dot or dot-density maps are used to display point data, whereas choropleth maps are used for areal data, and contour or isopleth maps are used for continuous surface data." (Rezaeian *et al.*, 2007). The production of health data maps is subject to several statistical constrains connected to the very nature of the data studied. *"For instance, mapping rates in small areas tend to create a misleading picture while using statistical significance, particularly in areas with large populations, produce small p values indicating statistical significance, but do not disclose scientifically interesting differences. The mapping of standardised rates is generally preferred to the mapping of p values, controlling for the influence of sampling variation by using a smoothing technique."* (Rezaeian *et al.*, 2007). The production of rates for small geographic units involves greater statistical variability (rate instability) and they become non interpretable. It is therefore difficult to produce a map on finer zones without violating these statistical constraints.

That is why the underutilization of certain types of maps (e.g. density maps (i.e. Smoothed maps)) or specific algorithms (e.g. Krigeage) is linked to the application field rather than a misunderstanding of the GIS tools or the complexity of using them on the part of professionals. And there is also the notion of data confidentiality that prevents the dissemination of data on geographic units that are too small. According to Elliot and Wartenber (2004) *"Confidentiality may also be an important issue. Breaching the confidentiality of spatial data may cause concern, especially when it discloses areas with high rates of morbidity/mortality or high levels of pollutants."*

However, (Rushton 2003; Feier 2000; Rezaeian *et al.*, 2007) identified the use of more enhanced spatial analyses such as complex spatial analyses (e.g. categorization *(clustering)*, self-correlation, neighbourhood analysis, vector generating and network analysis) in the health field. This proves that their usage can be more appropriate based on specific needs. For a complete understanding of the public health surveillance statistical methods, the reader is invited to consult the review published on the subject by Sonesson and Bock (2003).

Chung *et al.* (2004) also confirmed that GIS are not used for their statistical analysis functions, but rather for their geocoding and *buffer zone* capacities. He proposed the integration of statistical tools that were better adapted to the health field and GIS in order to facilitate the spatial aggregation of data. This recommendation was supported by Scotch *et al.* (2006), which concluded a real joint use of these tools among the 27 users interviewed in his survey on the use of GIS for the resolution of public health problems. The results indicated that most of the respondents used a statistical tool in conjunction with GIS. It therefore appears that the current capacities of GIS tools for statistical analysis are not sufficient for their usage in the health field and that their capacities could be increased in this field through the more efficient coupling with tools dedicated to data calculations.

Towards the end of the 20th century, the rapid growth of applications on various systems and the need to export the applications on the Internet required the development of data formats and transfer standards. Currently, GIS applications are most often deployed with web mapping technologies (see section 4.2), making it possible to offer access to data to a larger number of users. These applications are more and more developed on open source solutions (see technology, section 4.2.d), that operate on a full array of operating systems and which can be easily personalized. Furthermore, recent developments in terms of interoperability (e.g. WMS and WFS standards of the Open Geospatial Consortium) are greatly adopted in the field and have democratized the basic capacities of these systems and facilitate their implementation. Considering that in environmental health, data from various sources are frequently integrated, these new interoperability capacities represent undeniable benefits (increased ease in integration, possibility of connecting directly to the source, reduced updating problems, etc.).

4. 1. c. ***Existing Technologies***

Several GIS technologies exist on the market, such as CARIS[6], Geomatica[7] from PCI Geomatics, ArcGIS[8] from ESRI, Geomedia[9] from Intergraph and *MAPINFO*[10]. Other technologies that are more specific to the health field exist such as **EpiMap,** developed by the Center for Disease Control and Prevention (http://www.cdc.gov) and the World Health Organization (WHO), designed to be viewed in the form of data maps generated by EpiInfo. An equivalent technology, **SIGEpi**[11], was developed by the *Pan American Health Organization*, a regional office of the World Health Organization.

4. 1. d. ***Examples of Health Applications***

As mentioned in the introduction, the number of applications deployed in the health field in Canada and elsewhere in the world is significant. However, most existing systems are based on the transactional approach because they were developed and implemented many years ago using the technologies available at that time. For example, the World Health Organization distributes

---

[6] *Caris, http://www.caris.com*
[7] *PCI, http://www.pcigeomatics.com*
[8] *ESRI, http://www.esri.com*
[9] *Intergraph, http://www.intergraph.com*
[10] *Mapinfo, http://www.mapinfo.com*
[11] *SIGEpi http://www.paho.org/English/DD/AIS/sigepi_web2003en.htm*

the ***Health Mapper***[12], a GIS application under *MapObjects* by ESRI, which comes with its own database and spatial data sets.

In Canada, almost ten years have passed since the need to integrate GIS technologies in the Canadian health network has been expressed. Since then, *Canada's Health Infoway*[13] has been encouraging the development of computer networks (e.g. e-health), information technologies and web mapping for disease surveillance. The infoway supports the interoperability and utilization of standards among health information systems and includes a standards review process that take into consideration the standards and proposals of the OGC/CGDI. Also, the *GeoConnections* program of the Government of Canada is working in collaboration with the Canadian Institute for Health Information, the Public Health Agency of Canada and the Canada Health Infoway to facilitate the integration of health databases in the Canadian infrastructure and to date has funded the development of several applications and computer systems that will be deployed in the health network[14].

Lastly, several web mapping applications have been applied to various aspects of health surveillance in Canada. The Public Health Agency of Canada has developed a ***Disease Surveillance on-line website***[15] which provides mapping and various other data and services on cardiovascular disease, cancer, chronic diseases, etc. In 2001, the Public Health Agency of Canada implemented a national surveillance committee for West Nile Virus infections. Since then, several tools for West Nile Virus surveillance have been developed in Canada (Shuai *et al.*, 2006), as in certain provinces such as Quebec (Gosselin *et al.*, 2005) and British Columbia (British Columbia Center for Disease Control, 2007).

The following section therefore presents the unique characteristics of web mapping tools and their various applications.

---

[12] *Health Mapper, http://www.who.int/health_mapping/tools/healthmapper/en/index.html*
[13] *Infoway, http:// www.infoway-inforoute.ca*
[14] *These applications will be presented in greater detail in the next sections of the report.*
[15] *http://www.phac-aspc.gc.ca/dsol-smed/*

4.2 WEB MAPPING

4. 2. a. ***Summary Definition***

Web mapping can be defined as *"a group of products, standards and technologies that provide access to geographic information, in the form of maps, over the Internet."* (Web Mapping, Open GIS Consortium Glossary, 2007).

4. 2. b. ***State of the Art***

The democratization of geographic information that we are witnessing today is mainly attributable to the latest developments in the web. In fact, its sophistication enabled the growth of a completely new form of mapping: web mapping. From a quantitative point of view, the web has become the most important media for the dissemination of maps (Kraak *et al.*, 2001). According to Peterson (2003), the number of maps disseminated on the Internet in 2003 was estimated at more than 200 million per day. This interest is completely natural, given the many benefits offered by web mapping compared to paper maps. Notably, access is unparalleled, since the users can gain access at any time and from practically any location. Moreover, the maps dynamically created on the web are more interesting than paper maps because they automatically reflect data updates. Lastly, they offer interactivity that is non-existent with paper maps.

Obviously, the field of web mapping found its beginnings on the Internet, which appeared towards the end of the 1960s. Its visible part, the Web, saw the light of day in 1989, in the laboratories of the *European Organization for Nuclear Research (CERN)*. However, it would only be in 1993 that the first mapping server, launched by Xerox[16] was implemented. The following year, the Government of Canada launched the first version of its national atlas (Ferland, 2006). However, it was the arrival of the famous *MapQuest* in 1996 that would popularize the new trend of web sites with a geographic mapping flavour. According to Peterson (2003), it was in 1997 that the web experiences real growth in the dissemination of interactive maps.

In 1998, one of the first web services for the dissemination of aerial images, *TerraServer*, went online following a joint initiative by Microsoft, Hewlett Packard and the United States Geological Survey (USGS). More recently, in 2005, the very popular *Google Map, Google Earth* and *Virtual Earth* by Microsoft appeared.

[16] *http://www2.parc.com/istl/projects/www94/iisuwwwh.html*

Like the Internet, web mapping is evolving at breakneck speed. In the beginning, it was limited to the presentation of static maps, often stemming from the digitization of paper maps. These maps were presented in the form of images, such as GIF, JPEG and PNG. Such image formats are very limited in terms of visualization and offer practically no interactivity. For example, a zooming operation does nothing but enlarge the map without modifying the contents, which results in a degradation of the visual quality. Subsequent web mapping applications benefited from client/server architectures in order to generate maps on the server side and then to transmit them in the form of images on the client side. Although they were still static and not very interactive, these maps nevertheless had the advantage of being generated dynamically based on data sources. In order to offer more flexibility, new applications based on vector data were developed. Based on a database approach, these applications allow for greater interactivity with the users, enabling them to control certain map viewing parameters, notably the information layers to be displayed (e.g. road network, buildings) and the level of map detail. New formats were then used by these applications such as FLASH and SVG.

Today, apart from the static maps that are still present on the web, several other types of maps are offered by web mapping applications. Notably, there are animated maps, making it possible to represent the temporal evolution of a phenomenon (e.g. weather maps), distributed maps designed by combining data from various map servers, real time maps used for managing road traffic, open and reusable maps issued by web servers and capable of being integrated in another web page via API (Application Programming Interface) (e.g. Google Map API). More recently, certain web mapping applications make it possible to define maps in a collaborative manner, like Wikipedia (e.g. OpenStreetMap[17] and WikiMapia[18]) and to collaboratively produce spatial data mashups.

Although the web offers several advantages for the dissemination of maps, there are certain limitations. Among the most important are the transfer rate, the size as well as the resolution of screens, which is infinitely less than paper products (for more details, see (Arleth, 1999). Moreover, the much greater accessibility of web applications compared to office applications make them more difficult to design, given that it is almost impossible to foresee the configuration of the client applications and especially the level of user knowledge and expertise. Lastly, certain fields such as health are particularly sensitive to data confidentiality and strategies must be implemented in order to ensure the security of nominative and clinical data.

---

[17] *OpenStreetMap, http://www.openstreetmap.org/*

[18] *WikiMapia, http://www.wikimapia.org/*

4. 2. c. ***Positioning in Relation with Other Technologies***

Web mapping applications are far from offering the same level of spatial analysis as GIS tools. Most of them only have simple navigation functions such as zooming in and out and panning. Some offer a few analytical capacities used notably in route planning. Among the reasons explaining these limitations, there are the technical constraints that make it difficult to export GIS functionalities on the web and the public targeted by these types of applications. In fact, internet users are rarely experts in geomatics and do not have complex needs in terms of spatial analysis for example.

However, for more experienced Internet users, geomatics specialists or experienced users in the health field, it might be interesting to have access to more advanced GIS functions, without necessarily having to acquire proprietary GIS software, which is often costly. This need resulted in web-based GIS applications (i.e. *WebGIS*). These tools look like web mapping applications, but with particular emphasis on spatial analysis functions. According to Yang *et al.* (2005) *"A WebGIS (also known as web-based GIS and Internet GIS) denotes a type of Geographic Information System (GIS), whose client is implemented in a Web browser. [...] (It) focuses on how to allocate spatial data, both raster and vector (Goodchild, 1992), in a client-server-based Web platform, as well as on how to allocate functions to different system components in processing these data to satisfy users' needs."*

The interrelation between mapping, GIS tools and the Internet can be illustrated with the following figure. It therefore appears that web mapping stems from the conjunction of mapping on the Internet and that the GIS on the web are the result of a need for more complete applications deployed on the Internet.

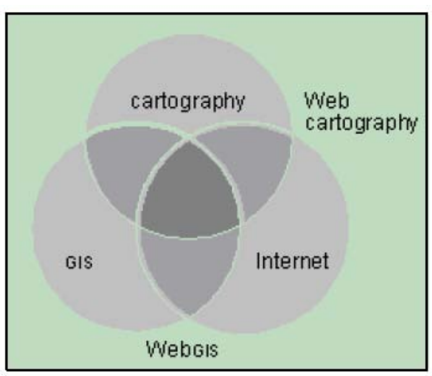


Figure 2. *Mapping, GIS and Internet*[19]

[19] *source http://www.webmapper.net/thesis*

4. 2. d. ***Existing Technologies***

Several commercial technologies exist in order to support the implementation of mapping/GIS applications on the web. Although they can be used in support of basic mapping applications, the web solutions proposed by the GIS publishers are mostly oriented towards the development of web-based GIS applications. Among these technologies are the following: *ArcIMS*[20] (ESRI), *GeoMedia WebMap*[21] (Intergraph), *MapXtreme*[22] (MapInfo), *Push'n'See*[23] (Korem), *JMap*[24] (KHEOPS), *Spatial Fusion*[25] (CARIS) and *MapGuide*[26] (Autodesk). Other commercial solutions are more specifically designed for the deployment of web-based mapping applications such as *Geoclip*[27], *GeoConcept*[28] and *AIGLE*[29] (Business Geografic). Moreover, several open source technologies have emerged over the last few years in the dissemination of geographic information on the Web such as PostGIS, *MAPresso*[30], *OPenMap*[31], *MapServer*[32], *Chameleon*[33] and *CartoWeb*[34].

4. 2. e. ***Examples of Health Applications***

Today, there is still an impressive quantity of web sites of a geographic character and those disseminating data associated with the health field have been emerging over the last few years. It is therefore necessary to categorize them in accordance with their specific characteristics. As stated previously, web mapping applications are far from offering the level of spatial analysis as GIS tools.

Web-based GIS offer access to more complex GIS functions. These tools place particular emphasis on the spatial analysis functions such as for example the analysis of corridors.

To this can be added a category of applications very present in the field of health: digital atlases. For a long time now, public health professionals have been using atlases to present statistical results on certain pathologies affecting the population (e.g. cancer) or determinants of the state of health (e.g. air quality) in the form of a collection of maps, tables and charts. For

---

[20] *http://www.esri.com/software/arcgis/arcims/index.html*
[21] *http://www.intergraph.com/gmwm/*
[22] *http://extranet.mapinfo.com/products/overview.cfm?productid=1849*
[23] *Push'n'See est une solution clé en main basée sur MapXtreme, http://www.pushnsee.com/fr/a_propos/solution.jsp*
[24] *http://www.kheops-tech.com/fr/jmap/index.jsp*
[25] *http://www.caris.com*
[26] *http://www.autodesk.fr/adsk/servlet/index?siteID=458335&id=7499959*
[27] *http://www.geoclip.fr/fr/p11_webmapping.php*
[28] *http://www.geoconcept.com/?6/Solution-Internet-Intranet-GeoConcept-Internet-Server-GCIS*
[29] *http://www.business-geografic.com/web/index.php?option=com_content&task=view&id=16&Itemid=1&lang=fr*
[30] *http://www.mapresso.com/*
[31] *http://openmap.bbn.com/*
[32] *http://mapserver.gis.umn.edu/*
[33] *http://chameleon.maptools.org/index.phtml*
[34] *http://cartoweb.org/*

example, the World Health Organization proposes the ***Atlas of Health in Europe***[35] in printed and downloadable version (in Acrobat) from its Internet site. However, Boulos (2004) identified a problem with these static atlases: *"Even if the WHO keeps publishing updated versions of this atlas, it will always lack (in its current form) the interactivity, real time or near-real time processing of current data, and the proactive features desirable in a true regional/community public health surveillance and spatial decision support system."* Consequently, more and more digital atlas applications are being developed. According to Bell *et al.* (2006) *"Changes in the production of atlases have also produced new analytic and communication opportunities. Historically, atlases were designed as books. However, over the last 10 years, mapping of health data has progressed from static maps designed for print media where the author selected both data and layout, to dynamic, interactive mapping over the Internet where the public may produce maps for their own purposes."* Again according to Bhowmick *et al.* (2006), *"Today, many health atlases are published online, presenting an opportunity for designers to make them interactive, animated, extensible and linked to other data display methods."* However, this type of application is in fact a specific use of GIS technology or web mapping, to which have been added data viewing components such as tables, charts and diagrams.

These digital atlases cannot be categorized as analytical applications, because although they target similar objectives, they do not make use of technologies from this field (*cf.* section 3.0), or the multidimensional approach, which is the basis of business intelligence. However, it can easily be stated that this type of application is the transactional ancestor of dashboards (see section 5.3) in the analytical world. The digital atlas solidly reflects the need for analytical tools supporting decision-making in the health field. However, the health professionals who work in mapping or GIS who are not perhaps sufficiently informed on the technological possibilities of the analytical world, they develop decision-making applications using transactional technologies which they understand better. Conversely, health professionals who use decision-support technologies are not sufficiently informed on the technological possibilities of the world of geomatics. This observation is the general rule currently in all fields.

There are commercial products that enable the development of digital atlases such as *InstantAtlas* and *Instant Profiler*[36] by GeoWise, which make it possible to produce maps using vector spatial data in ShapeFile format (ESRI) or MID/MIF format (MapInfo). No mapping server is necessary. However, the spatial data is disseminated in SVG format (Scalable Vector Graphics). The SVG Viewer freeware is therefore necessary to consult the maps. The GIS

---

[35] *Atlas of Health in Europe, http://www.euro.who.int/document/E79876.pdf*
[36] *InstantAtlas and Instant Profiler by GeoWise http://www.instantatlas.com/*

functions in the application are however limited to zoom, pan, resize and the selection of mapped elements.

i. Examples of Web Mapping Applications

The Americans, who are required to publicly disseminate government information, are large producers of web mapping applications. The application ***Web Interactive Cancer Mortality Maps***[37] developed by the National Cancer Institute (NCI) and the National Institutes of Health (NIH) is a good example of this and presents the geographic patterns and temporal tendencies of cancer mortality rates (for more than 40 cancer sites) during the 1950 to 1994 period. For its part, the application ***CDC Injury Center's Interactive Mapping System***[38] provides access to the geographic distribution of mortality rates by injuries in the US.

The ***USGS National Atlas of the United States***[39] proposes an interactive web mapping environment and offers several layers of data from a multitude of fields, including health. It includes notably data on several causes of mortality and West Nile Virus infections. It enables the superposition of health data with environmental information. The navigation and querying functions are minimal however.

Based on the Geoclip technology, the ***Montreal Health Atlas***[40] is a web mapping application that makes it possible to disseminate information on the population and the health system in the Montreal region. The application presents notably the socio-demographic data and the use of health services by the population based on various geographic sections. This section is completed by a ***strategic dashboard***[41] which presents the indicators that make it possible to follow the evolution of the priority transformation in the health and social services network in Montreal (see page 65).

Currently, funding by the Government of Canada's *GeoConnections* program will enable the deployment of several GIS applications on health surveillance. The software and technology used as part of these projects are compliant with the interoperability standards established by the Government of Canada and will be part of the global Canadian Geospatial Data Infrastructure (CGDI). The CGDI renders the geospatial data accessible to all Canadians through web mapping services.

---

[37] *http://www3.cancer.gov/atlasplus/index.html*
[38] *http://www.cdc.gov/ncipc/maps/*
[39] *http://www.nationalatlas.gov/natlas/Natlasstart.asp*
[40] *http://www.cmis.mtl.rtss.qc.ca/fr/atlas/atlas_presentation.html*
[41] *Strategic dashboard http://www.cmis.mtl.rtss.qc.ca/fr/performance/tableaudebord/tb_presentation.html*

First of all, in 2003, the Public Health Agency of Canada put online an ***interactive GIS application for the surveillance of the West Nile Virus***[42] using MapServer, an open source web mapping application development environment. This project makes it possible to notify in almost real time the human cases diagnosed as part of West Nile Virus surveillance. Health Canada, in collaboration with *GeoConnections*, has amalgamated its various databases so as to analyze the integrated data on the West Nile Virus and to increase its capacity to retrieve geospatial information, access data, manage information and business-intelligence processes.

Another piece of funding granted by GeoConnections will enable the development of a ***Health Surveillance Map Interface***[43]. More specifically, the project will allow the EpiInfo map component (an internationally tested software) to access Health Canada's spatial data warehouse. Using the simple map interface, health professionals can now access national spatial data using a web browser, extract the information desired (hospital locations, levels of income, cancer rates or any other parameter to be analyzed) and add the data to the maps. This project uses technology developed by *CubeWerx*[44] in Gatineau.

From a global point of view, the World Health Organization (WHO), through the *Public Health Mapping and GIS program* has been developing, since 1993, geographic tools and applications in order to support decision-making regarding infectious diseases and public health programs. Notably, the WHO has developed the first ***online world atlas on infectious diseases***[45], a new tool for the surveillance and control of diseases based on more than 300 indicators for various geographic levels. The analysis and interpretation of data are supported by complementary information of a demographic, socioeconomic and environmental nature. For report needs, the data can also be disseminated in the form of charts and tables. Moreover, the WHO is also responsible for the ***Health Mapper*** application, a GIS tool specifically designed for users in the public health field (see section 4.1).

ii. Examples of GIS Applications on the Web

In 2004, the New Brunswick Lung Association[46] developed an ***electronic mapping application***[47], which facilitates decision-making in the fields of environmental and health protection. This web mapping technology favours the sharing of information and access to data on the Internet, displayed based on geographic location and offers the user an uninterrupted view

---

[42] *http://www.cnphi-wnv.ca/human2006/index_f.htm*
[43] *http://www.geoconnexions.org/ICDG.cfm/fuseaction/projects.projectDetails/id/32/gcs.cfm*
[44] *http://www.cubewerx.com*
[45] *http://globalatlas.who.int/*
[46] *http://www.nb.poumon.ca/html_fr/Programmes/Sante_pulmonaire/cartographie.htm*
[47] *http://www.geoconnexions.org/ICDG.cfm/fuseaction/successStories.seeFile/id/1122/gcs.cfm*

of the information. The application is based on the web mapping technology *Spatial Fusion* by CARIS. A more recent version of the application, developed in 2006, offers more possibilities through the adding of temporal viewing functionalities through ad hoc developments making use of OGC standards in compliance with general expectations. These efforts are an example of the necessity to improve transactional technologies in order to achieve certain decision-making objectives.

Several Canadian provinces have followed suit in the development of surveillance applications for health and infections due to the West Nile Virus such as the British Columbia Centre for Disease Control, which has developed the ***Interactive GIS Mapping for West Nile Virus***[48]. It makes it possible to map dead bird cases, analyzed mosquito batches and cases of human infection. This web-based GIS application, developed using *ArcIMS* by ESRI, uses an HTML viewer. These functions are much more advanced than the web mapping application used by the Public Health Agency of Canada since it requires the use of a GIS.

In order to support the decision-making by health professionals assigned to the West Nile Virus file (WNV), the Institut national de santé publique du Québec (INSPQ) was mandated in 2003 by the ministère de la Santé et des Services sociaux to develop a ***Système intégré des données de vigie sanitaire du virus du Nil occidental [Integrated System for Health Surveillance Data on the West Nile Virus] (SIDVS-VNO)*** (Gosselin *et al.*, 2005; INSPQ, 2004), which is capable of both grouping surveillance data together from several sources and making them available in real time. The SIDVS-VNO facilitated the gathering, localization, management and analysis of data. It also facilitated the analysis of results through maps, tables and statistical diagrams. It is a system that uses the *JMap* mapping server and a *Microsoft SQL Server* database.

[48] *Application from the BC Center for Disease Control: http://maps.bccdc.org*

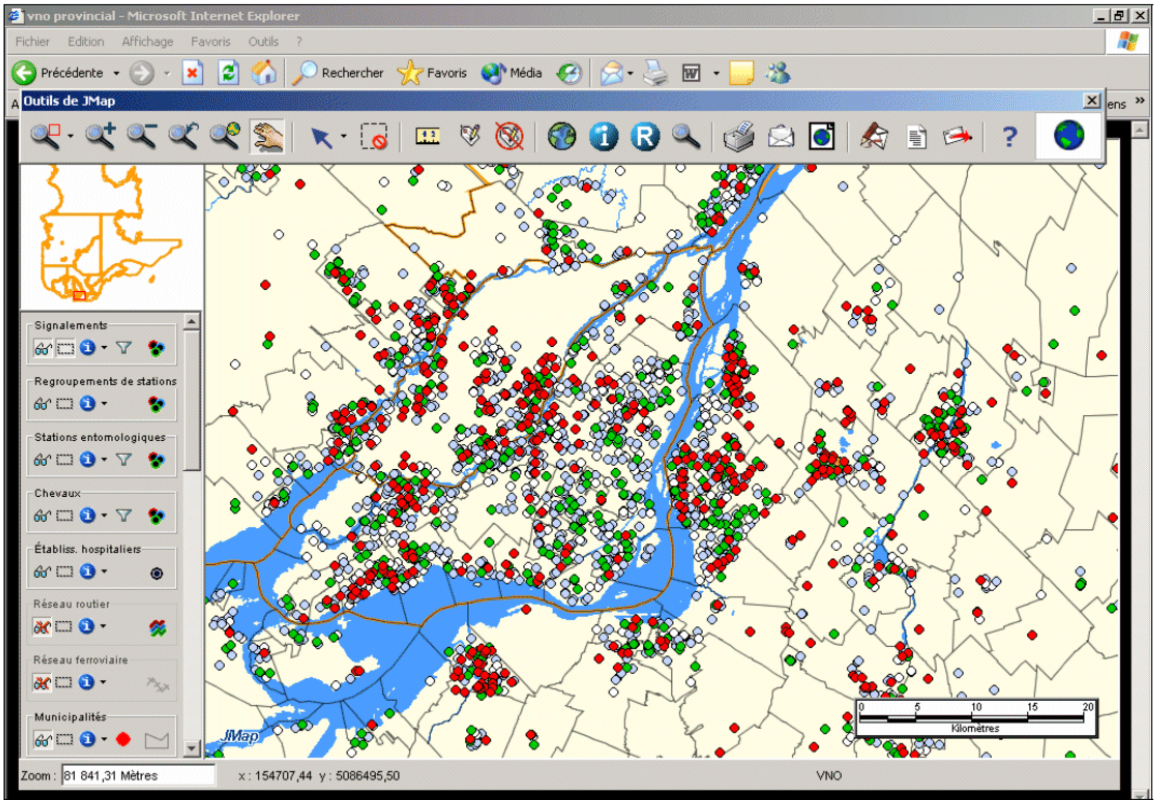


Figure 2. Example of corvidae reporting presented in the
Système intégré des données de vigie sanitaire du virus du Nil occidental (SIDVS-VNO).

iii. Examples of <u>Digital Atlas</u> Applications

An array of application examples is available on the *Instant Atlas* site[49], a tool dedicated to the creation of digital atlases. Our searches on the web and in scientific applications also made it possible to identify several applications from the health field in general. In France, the Observatoire régional de la santé en Provence-Alpes-Côte d'Azur has developed the ***Système d'information régional en santé [Regional Health Information System] (SIRSé)***[50]. SIRSé makes available to everyone all the information on the health status of the regional population, at the territorial level. It makes it possible to easily access data in the form of interactive maps and tables. For each indicator, it is possible to access its definition, information about its data sources, the methodology and the precautions to be taken in interpreting the results. The application was developed using the *Geoclip* software. Figure 4 presents the interactive map for the comparative rates of breast cancer mortalities in women by nearby territories, for the 1995-1999 periods.

---

[49] *http://www.instantatlas.com/health.xhtml*
[50] *SIRSé : http://www.sirsepaca.org/selection_indicateur.php*

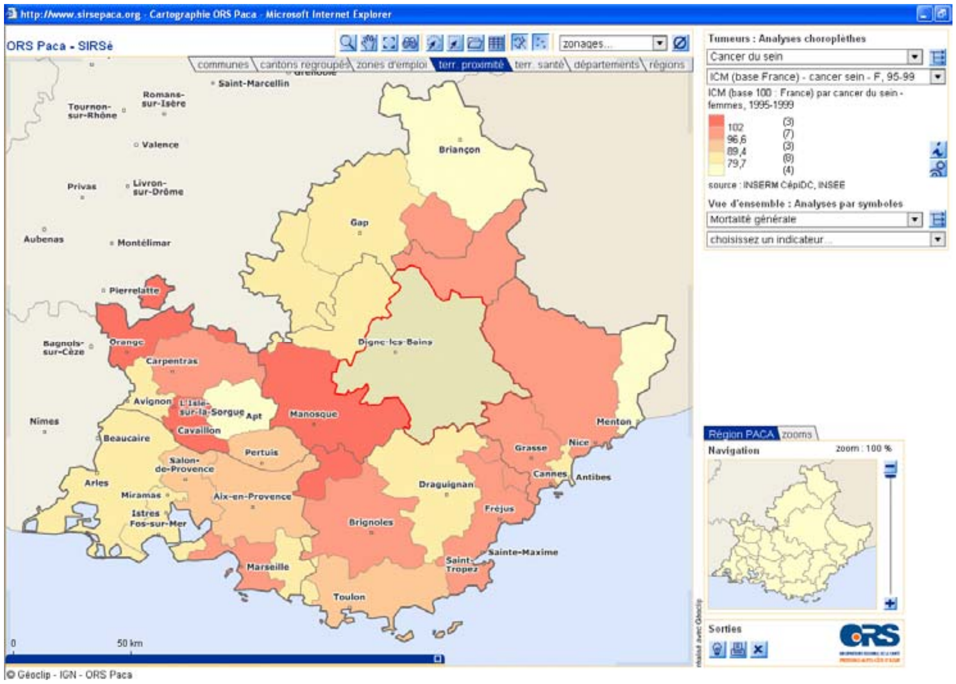


Figure 3. Interactive SIRSé map presenting the comparative rates of breast cancer mortalities in women by nearby territories, for the 1995-1999 periods.

In the US, the ***Pennsylvania Cancer Atlas***[51] produced by the GeoVista Center, at Pennsylvania State University, is an interactive online tool dedicated to epidemiologists, professionals, politicians and managers (Robinson *et al.*, 2005). This interactive web tool is composed of four displays presenting the incidence rates for various cancers over a long period of time. The first three (map, table and diagram) are synchronized to present the cancer incidence rate by county. Added to this is a histogram of the population breakdown by gender and race. Clicking on a country of the map makes it possible to produce the histogram corresponding to the population of this region. It is also possible to open two maps in order to conduct temporal or spatial comparisons. The application was developed on a *PostGIS* server and the data is projected on a *MacroMedia Flash* client. Figure 5 presents the atlas's dashboard for the incidence of colorectal cancer by county, for the 1994-2002 periods.

---

[51] *Pennsylvania Cancer Atlas http://www.geovista.psu.edu/grants/CDC/*

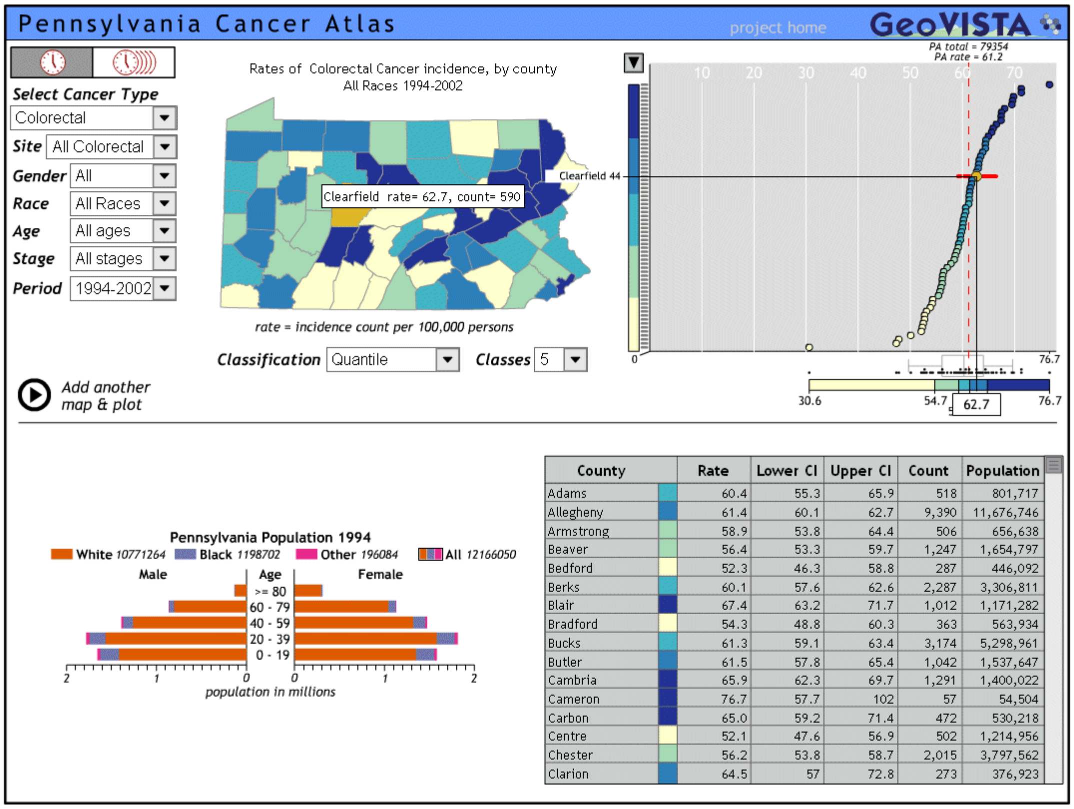


Figure 4. Pennsylvania Cancer Atlas presenting the incidence of colorectal cancer by county, for all races, for the 1994-2002 periods.

Several other US applications of the same type are presented on the web, e.g. the ***Heart Disease and Stroke Maps Interactive State Maps***[52] (National Center for Chronic Disease Prevention and Health Promotion), the ***Kentucky Cancer Registry: Interactive Cancer Mapping***[53] (University of Kentucky), ***EpiQMS***[54] (Department of Health) and the ***State Cancer Profiles Interactive Maps***[55] (National Cancer Institute).

In Canada, the Public Health Agency of Canada supports an ***infobase for the surveillance of non-communicable diseases (NCD)***[56], which makes it possible to establish the epidemiological profile of the principal non-communicable diseases in Canada, notably the more current cancers and cardiovascular and respiratory diseases by province or territory and by regional health service. The NCD Surveillance Infobase (also called the CVDInfobase) uses a *Visual Basic* application using *MapObjects* by ESRI as a GIS component. Several display options are available, i.e. comparative data between a large number of regions, temporal trends pertaining

---

[52] *Heart disease and stroke Interactive State Maps, http://apps.nccd.cdc.gov/giscvh/map.aspx*
[53] *Kentucky cancer registry : http://www.kcr.uky.edu/*
[54] *EPIQMS, http://app2.health.state.pa.us/epiqms/Asp/ChooseDataset.asp*
[55] *State cancer profiles, http://statecancerprofiles.cancer.gov/map/map.noimage.php*
[56] *Infobase, http://www.cvdinfobase.ca/surveillance/Mapdb/Infobase_f.htm*

to morbidity and mortality, mortality trends based on a cohort of births and proportional mortality trends. This interactive dashboard allows for the successive display of three viewing modes (map, table and chart). Please note that the Canadian equivalent, supported by the World Health Organisation (WHO), exists under the name ***Global Cardiovascular Infobase***[57]. Figure 6 illustrates a map of breast cancer mortality rates in Quebec.

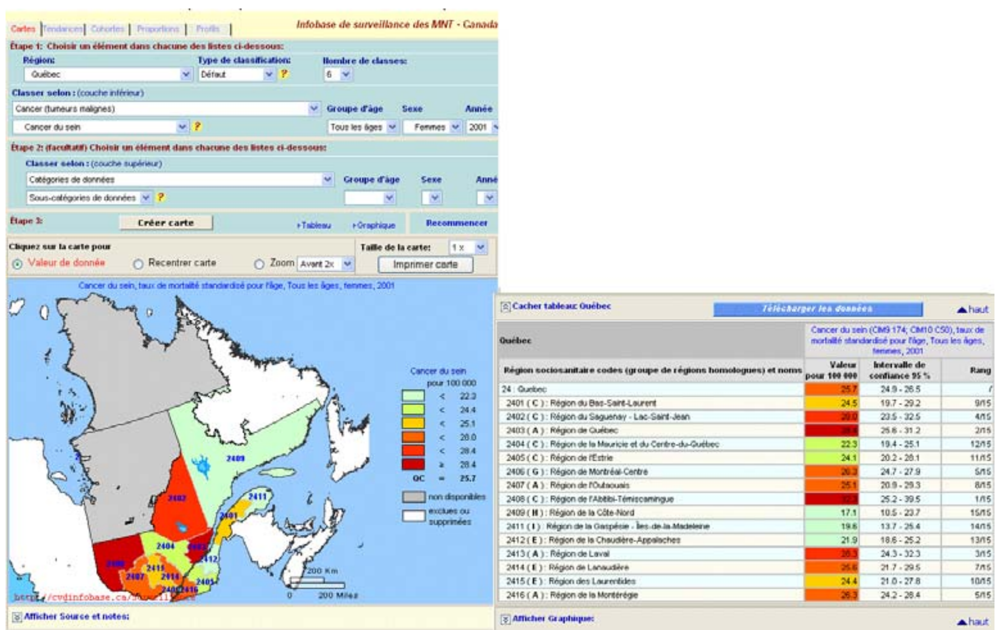


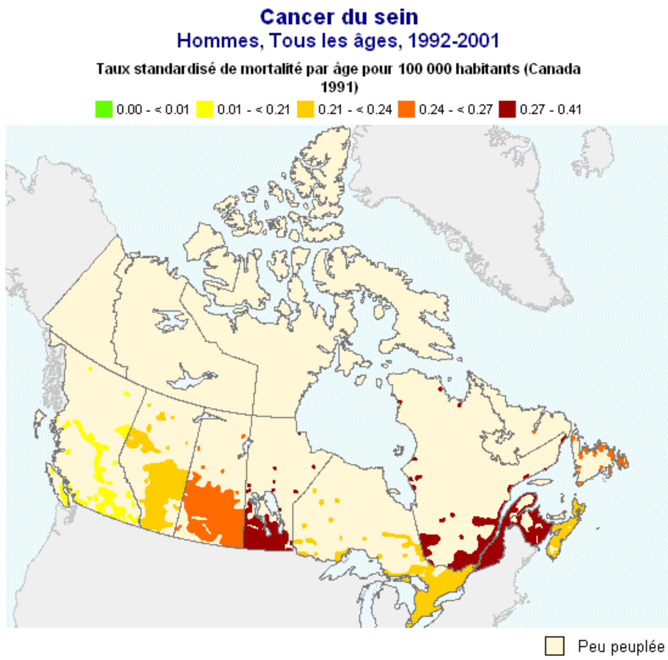
Figure 5. Breast cancer mortality rate map for Quebec along with the table of associated data

Disease Surveillance On-Line[58] by the Public Health Agency of Canada is a tool that offers four applications to the web site's users allowing them access to data on cancer, cardiovascular disease and notifiable diseases for various years and allows them adapt this data to their specific needs. Like the previous application, this interactive dashboard allows for the successive display of three viewing modes (map, table and chart). Figure 7 illustrates a map of the breast cancer mortality rates in Canada.

---

[57] *Global cardiovascular infobase http://www.cvdinfobase.ca/*

[58] *Disease Surveillance On-Line http://www.phac-aspc.gc.ca/dsol-smed*



Obtenir valeur d'emplacement

Province	Taux/100 000	Limite inférieure	Limite supérieure
Terre-Neuve	0.26	0.10	0.55
Île-du-Prince-Édouard			
Nouvelle-Écosse	0.21	0.10	0.37
Nouveau-Brunswick	0.40	0.23	0.66
Québec	0.41	0.34	0.49
Ontario	0.24	0.20	0.28
Manitoba	0.27	0.16	0.44
Saskatchewan	0.24	0.13	0.40
Alberta	0.23	0.16	0.34
Colombie-Britannique	0.20	0.14	0.27
Yukon			
Territoires du Nord-Ouest			
Nunavut			

Figure 6. Map of breast cancer mortality rates in Canada along with the table of associated data

5. Analytical systems

In order to optimize the analysis, analytical systems generally rely on a completely different paradigm than that of transactional systems, i.e. the multidimensional paradigm as defined in the Business Intelligence and Data warehousing world (i.e. data cubes). The structures based on this paradigm have demonstrated, over the last 30 years, that they are easier to understand by users than transactional structures and are therefore easier to query (Kimball, 2002b). The multidimensional structures are used to store warehouse data and data mart data in a more adequate format for complex analyses. Before presenting the concepts pertaining to analytical systems, the underlying multidimensional structure will be presented.

The multidimensional approach introduces new concepts that until now have been unknown in the transactional field such as "dimensions", "measures", "facts" and "cubes". The *dimensions* are analysis themes (e.g. time, disease, territory). A dimension is composed of "members" that are organized in accordance with a certain hierarchy based on the level of detail of the members (e.g. year, month, day for a temporal dimension). The members of the lower level (e.g. the days) are aggregated together to compose the members of the higher detail level in the hierarchy (e.g. the months). The *measures* are the number values analyzed for the various dimensions (e.g. cancer rates). A set of aggregated measures based on a set of dimension form a data cube. Within the data cube, the various combinations of dimensions and measures form the "facts" (e.g. the incidence rate (measure) for cancer A (disease dimension) for region 1 (territory dimension) in 2000 (time dimension) was 62.7 per 100,000 person-years). These facts are grouped together in the fact table (described below) to form the analysis field. The techniques used to implement the multidimensional models are the star, snowflake or mixed schemas (Kimball, 2002a). The fact table is the central point of the model. The symmetry for such a model makes it so that each dimension has an equivalent entry point in the fact table. The response time for queries, regardless of their complexity, can therefore be predicted (Kimball, 1997). Furthermore, the changes in user query habits will not affect the model, since all queries are conducted in the same way.

The multidimensional structure considers the attributes like a dimension, a bit like spreadsheets, and each piece of information is deducted from the intersection of these dimensions. Two dimensions can easily be viewed in a matrix, but it is more difficult to imagine a three-dimensional structure (cube) or a structure with even more dimensions (hypercube). Figure 8 illustrates the concept of a multidimensional cube and represents the incidence rate based on *territory*, *time* and *diseases* dimensions. The data segment (see item A) represents the incidence rate for breast cancer for all the years and for all regions. The data cell (see item B) at the bottom represents the incidence rate for lung cancer in region C for the year 2002.

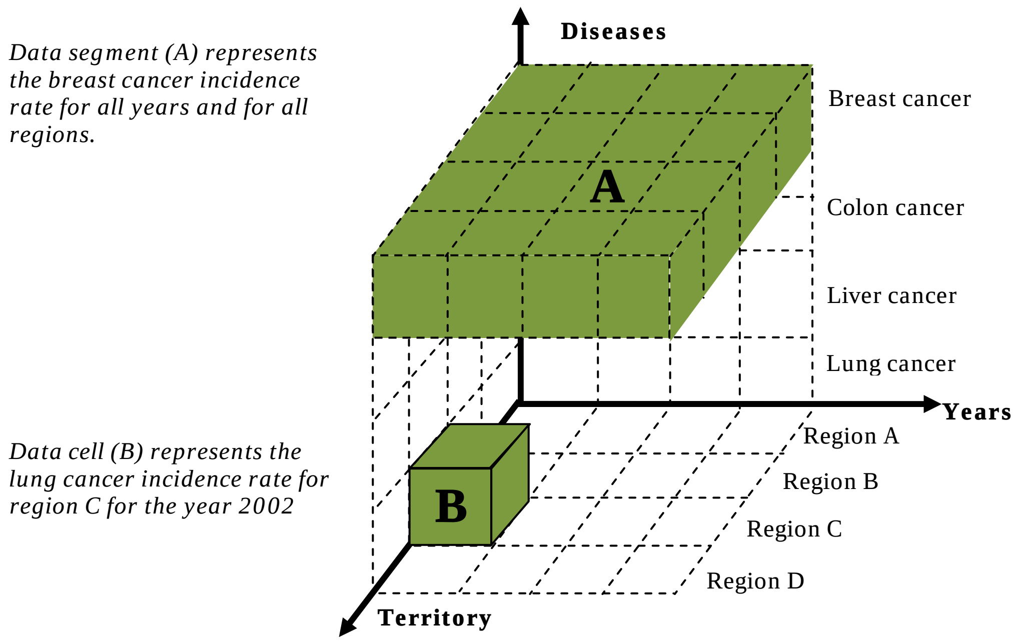


Figure 7. Illustration of a multidimensional cube representing the incidence rate based on territory, time and diseases.

Although it is not mandatory to implement a data warehouse to take advantage of the benefits associated with this data cube paradigm, such warehouses are frequently seen to help in the decision-making for an organization as a whole. However, even if all the warehouses do not necessarily have a multidimensional structure, it is when they adopt one that they take on their full value.

5.1 SPATIAL DATA WAREHOUSES

5. 1. a. *Summary Definitions*

*"A **data warehouse** is a single, complete and coherent deposit of data obtained from a variety of sources and accessible to users in a manner that allows them to understand this data and to use it in a business context."* (*Devlin 1997).*

*"A **data warehouse** is a subject-oriented, integrated, time variant and non-volatile collection of data in support of management's decisions."* (Inmon, 1996a).

*"A **geospatial data warehouse** is a collection of subject oriented spatial data that is of known quality, is non-volatile, is time variant and includes the basic tools to access and extract information."* (Rawling *et al.*, 1997).

5. 1. b. *State of the Art*

i. Concepts of the Non-Spatial World

According to Inmon (1996a), the concept of warehouse was born of the desire to integrate heterogeneous data disseminated in the various transactional systems of an organization. This concept was put forth in the mid 1980s (Inmon, 1996a; Devlin, 1997) and has become very popular in the last decade. A data warehouse is designed using:

- Data from fields affecting an organization as a whole;
- Non-volatile data (i.e. no updating is allowed, only additions);
- Integrated data from the systems in place;
- Data varying in time (i.e. current and historical data);
- Data at various levels of detail that is used to support the decision-making processes.

When designing a data warehouse, the data from transactional systems are copied, transformed and integrated within the same analytical structure. They undergo various operations that are grouped together under the term ETL (*Extraction, Transformation and Loading*). These operations are recognized as being the most time-consuming task in the implementation of a data warehouse. Within this context, the source transactional systems are often called *legacy systems* or operational systems and continue to execute the tasks that they have always conducted. Quite simply, the warehouse provides an added value to source data by integrating a copy of it in a new homogenous view to provide summary, aggregate, temporal, comparative information... i.e. analytical support.

Whereas the data warehouse is implemented so as to offer a global vision of a business, by including a copy of all the information of interest that are associated to it, the data mart targets needs specific to an activity sector of an organization. As a result, it can be an exact or modified subset of the data warehouse (Bédard *et al.*, 1997). According to Gill, *"A data mart is an implementation of a data warehouse with a small and more restricted scope of data and data warehouse functions serving a single department or part of an organization"* (Gill *et al.*, 1996). It happens that data marts are created directly from data sources, which is the case when a warehouse had not previously been developed.

In order to properly summarize the various concepts presented to date, figure 9 identifies the terminology to be used for each system based on the type of analysis (OLTP or OLAP), the level of data integration (local or business) and the level of data detail (detailed or summary).

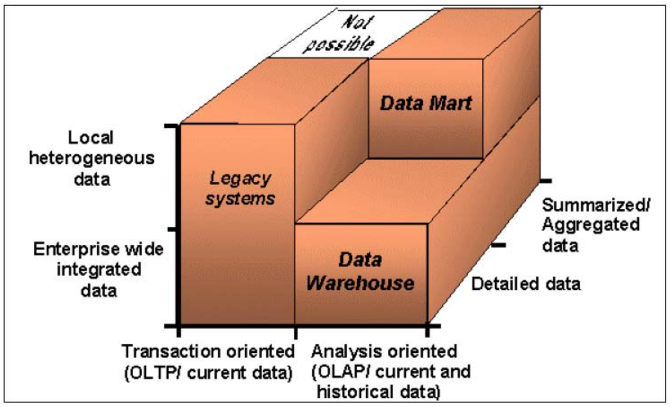


Figure 8. Positioning of transactional systems, data warehouses and data marts (Bédard et al., 1997).

In reading figure 9, it can be seen that the operational systems (legacy systems) only support detailed data; the *data warehouse* can support the aggregate and/or detailed data and the *data mart* supports the aggregated data integrated locally. As a summary, the following table compares the *operational systems*, the *data warehouses* as well as the *data marts*.

Transactional Systems	**Analytical Systems**	
Operational System	Data Warehouse	Data Mart
Focussed on transactions	Focussed on analysis	Focussed on analysis
Detailed data	Detailed and aggregate data	Mostly highly aggregate data
Integrated according to applications	Integrated for the enterprise	Integrated by subject or department
Continually updated	Never updated, addition only of new data	Never updated, addition only of new data
Current data	Current, historical and forecasting data	Current, historical and forecasting data
Original data source	Data imported from the transactional systems	Data imported from transactional systems and/or the warehouse
Normalized structure	De-normalized structure	De-normalized structure

Table 1. Comparison between the operational systems, data warehouses and data marts (Bédard et al., 1997).

The implementation of a data warehouse can be done via various architecture configurations. The most common is the centralized architecture (see Figure 8) where the warehouse accessed by client tools takes place at the centre of the architecture.

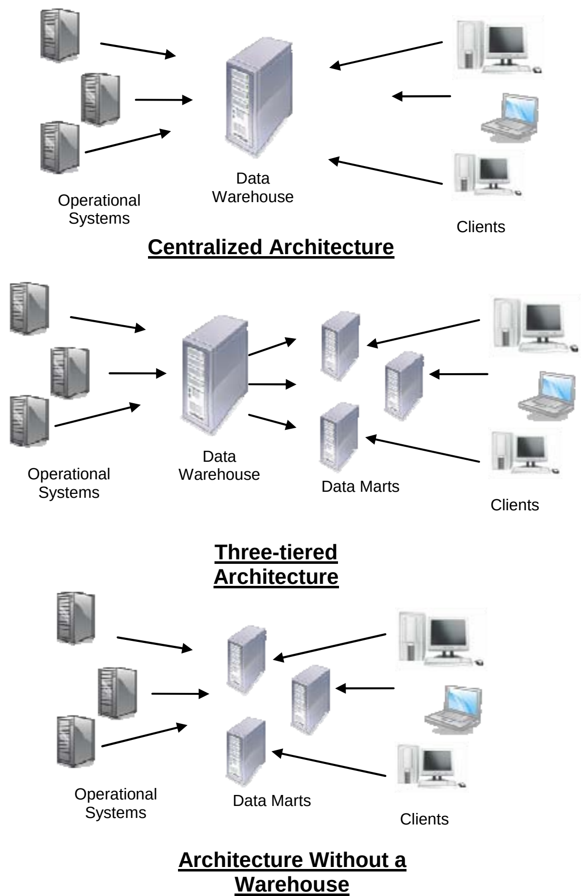


Figure 9. Data Warehouse Architectures

Another more complex architecture, the three-tiered architecture, allows for the deployment of the warehouse and various data marts constructed based on the subjects of the warehouse. Various architecture variants are possible, including the implementation of a data mart without first deploying a warehouse, which is a frequent solution aimed a reducing the efforts for implementation in a subset only of the business and thereby accelerating the delivery timeframes (Radding, 1995). Consequently, it is not necessary to have recourse to other architectures to deploy decision-support solutions in the businesses. An analytical tool can even access a multidimensional structure constructed on a single operational system and respond adequately to decision-making.

Going back to figure 1 on the deployment of a corporate architecture dedicated to decision-making, let us add examples of warehouse architectures: a centralized architecture, a three-tiered architecture (with data marts), and an architecture without a warehouse (data marts only).

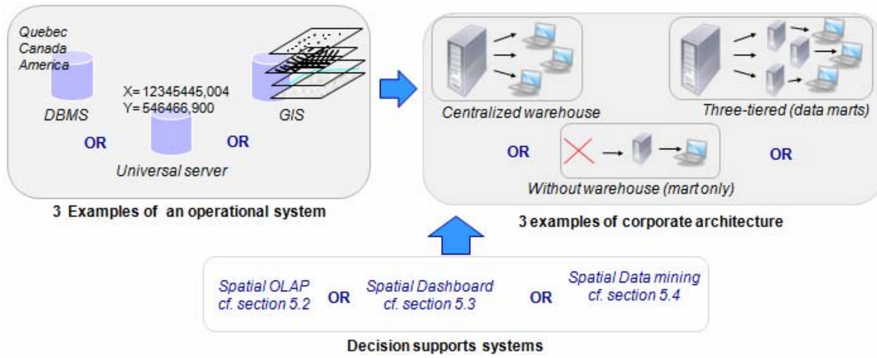


Figure 10. Positioning of data warehouse architecture examples in the deployment of a corporate architecture dedicated to decision-making

Lastly, specialized tools must be used in order to effectively explore and analyze the data stored in the analytical databases. These tools can be grouped in three principal categories: *querying and reporting tools*, *OLAP tools* and *data mining tools*. The first category of tools mainly supports the generating of reports and charts. The second category of tools, the OLAP tools, provides a more flexible exploration and analysis interface allowing for interactive knowledge discovery. Lastly, the third category is based on mathematical and artificial intelligence techniques in order to explore, in a fully automatic manner, warehouse data in order to identify trends, correlations, etc. To these categories are added dashboards that make it possible to combine various decision-support and transactional technologies in order to offer a decision-making interface to users. These various tools are respectively the subject of sections 5.2 (Spatial OLAP), 5.3 (Spatial Dashboards) and 5.4 (Spatial Data Mining).

ii. The Concepts of the Spatial World

The geographical nature of most data stored in a data warehouse has led to the design of warehouses capable of storing and managing the spatial component of data. *"It has been estimated that about 80 percent of all data stored in corporate databases are spatial data." (Franklin & Hane, 1992).*

The specific management of the spatial component in data warehouses then allows specialized tools (e.g. Spatial OLAP tools) to make use of this component in order to offer greater analytical capabilities. The presentation of analytical spatial and non-spatial tools is the subject of the next sections.

The spatial component brings with it new multidimensional concepts. Therefore, added to the descriptive dimensions found in the non-spatial data warehouses are three new types of spatial dimensions: descriptive spatial dimensions, geometric spatial dimensions and mixed spatial dimensions (see figure 12) (Han *et al.*, 1998; Bédard *et al.*, 2001). A spatial dimension is *"A type of dimension that associates a spatial reference to its members. This spatial reference is nominal in the case of a non-geometric spatial dimension, geometric in the case of a geometric* *spatial dimension, and a combination of both in the case of a mixed spatial dimension.*" (Bédard *et al.*, 2001).

1. Descriptive spatial dimension: the members of the dimension have a nominal spatial reference only (location name for example).
2. Geometric spatial dimension: the members of the dimension have a geometric component and can therefore be mapped.
3. Mixed spatial dimension: the members of certain levels have a geometric component, the members of other levels have a nominal spatial reference only.

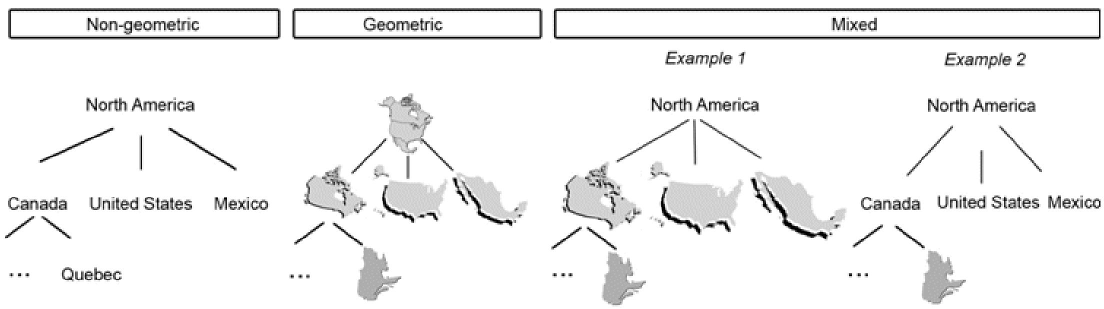


Figure 11. The Three Types of Spatial Dimensions (source: Rivest et al., 2003)

Like the dimensions, the measures from a spatial data warehouse can also possess a geometric spatial component (Han *et al.*, 1998). A spatial measure is "*A type of measure that can be obtained using spatial operators leading to geometric spatial measures (p. ex. polygon overlay) or numeric spatial measures (e.g. distance).*" (Bédard *et al.*, 2007). For example, a spatial measure can be the geometric result from the combination of spatial geometric dimensions (Bédard *et al.*, 2005). Therefore, the grouping of aggregate polygons resulting from the combination of forest stands (from a forestry dimension) and sub-basins (stemming from watersheds) is an example of a spatial measure (Rivest, 2000). Together, these elements combine to construct a spatial data cube that can be defined as "*A set of measures (spatial and non-spatial) organized according to a set of dimensions (spatial and non-spatial).*" (Bédard *et al.*, 2007).

5. 1. c. ***Existing Technologies***

At the present time, there are several solutions specially designed for the deployment of non-spatial data warehouses (e.g. *Microsoft SQL Server BI*[59], *TeraData*[60], *SAS Enterprise BI Server*[61], *InterSystems Caché*[62]). Some solutions explicitly manage the spatial component:

---

[59] *http://www.microsoft.com/sql/solutions/bi/default.mspx*

- *Oracle Business Intelligence,* based on Oracle 10g[63];
- IBM *DB2 Data Warehouse Enterprise Edition*, based on DB2 Spatial Extender[64];
- IBM *Informix Red Brick Warehouse* and Spatial DataBlade[65];
- CubeWerx Solutions by Cubewerx[66].

Furthermore, certain open source DBMS have a spatial component and can be used in the deployment of spatial data marts:

- *PostGIS* based on postgreSQL;
- *MySQL*[67] and its spatial extension;
- *MsSQLSpatial*[68] spatial extension by Microsoft *SQL Server*.

It is possible to combine certain traditional data warehouse solutions with spatial libraries in order to offer in-house solutions. For example, Microsoft *SQL Server* can be coupled with MapInfo *SpatialWare* technology[69] just like the *Essbase* technology by Hyperion can be combined with the *MapObjects* library from ESRI.

5. 1. d. ***Examples of Health Applications***

The research team at the *Colleges of Public Health and Business Administration* at the *University of South Florida* developed the ***Catch*** system (Comprehensive Assessment for Tracking Community Health) based on a data warehousing approach. This warehouse integrates data from the health field from several transactional systems in the State of Florida. The system allows health professionals to analyze more than 250 health indicators in the form of a report and is equipped with OLAP functions for enhanced analyses (Berndt *et al.*, 2001; 2003). However, no geographic data is supported in this system.

The *National Cancer Institute* (NCI), in partnership with the *National Institute of Environmental Health Sciences* (NIEHS), was mandated in 1993 to study the very high breast cancer rates on Long Island. This resulted in the creation of the *Long Island Breast Cancer Study Project*[70] (LIBCSP), which included an investigation into environmental factors, potentially

---

[60] *http://www.teradata.com/t/*
[61] *http://www.sas.com/technologies/bi/*
[62] *http://www.intersystems.com/cache/index.html*
[63] *http://www.oracle.com/technology/tech/bi/index.html*
[64] *http://www-306.ibm.com/software/data/db2/dwe/*
[65] *http://www-306.ibm.com/software/data/informix/redbrick/*
[66] *http://www.cubewerx.com/main/products/CubeSTOR.html*
[67] *http://www.mysql.com/*
[68] *http://www.codeplex.com/MsSqlSpatial*
[69] *http://dev.hyperion.com/resource_library/tips_faqs/kpis_business_analytics/essbase_geographic.cfm*
[70] *http://www.healthgis-li.com/default.jsp*

responsible for the high level of breast cancer in this region. To support this investigation, the GIS tool ***System for Breast Cancer Studies on Long Island*** (LI GIS) was developed. This system integrates more than 80 data sources (topographical, demographic, health, environmental, etc) within a spatial data warehouse (the ***LI GIS Data Warehouse***). The system is based on *ArcGIS* by ESRI, *Oracle* and *SAS*.

In Canada, the surveillance coordination centre at the *Public Health Agency of Canada* provides access to its spatial data warehouse to the Canadian public health community, as well as two tools, the ***Map and Data Exchange*** and the ***Public Health Map Generator*** (see GIS brochure for public Health practice[71] from the PHAC). The forum for the Map and Data Exchange makes it possible to collaborate and to exchange data and expertise between health professionals. Since July 2005, more than 200 professionals from 90 public health groups across Canada use this network. The Public Health Map Generator is an online tool that uses geographic information from the Canadian Geospatial Data Infrastructure (CGDI), which has a virtual warehouse of geographic data from a connected network of data suppliers across the country. For its part, the *Cancer Care Ontario* has deployed, since June 2006, the second implementation of the data warehouse on the web and the analytical tool ***iPort***[72] (TDWI, 2006). Lastly, the *Canadian Institute for Health Information (CIHI)* centralizes a record of key health indicators in its data warehouse and integrates a series of personalized reports on Internet called ***eCHAP.*** The first series of CIHI reports presenting the comparison of programs from hospital activities was deployed using *Discoverer* by *Oracle*. The application's Internet site supposedly offers access to mapping and geographic information, but nothing was identified as such when visiting the site in February 2007.

---

[71] *brochure, http://www.phac-aspc.gc.ca/php-psp/pdf/2005_brochure_gis_e.pdf*
[72] *http://cancercare.on.ca/index_requestDatafromCCO.htm#how_do_i_request_data*

5.2 SPATIAL ON-LINE ANALYTICAL PROCESSING (OLAP)

5. 2. a. *Summary Definitions*

"***On-line analytical processing*** *technology is a category of software focussed on the exploration and rapid analysis of data based on a multidimensional approach with several aggregation levels.*" (Caron, 1997).

"Spatial OLAP *is a generic software that allows rapid and easy navigation within spatial datacubes for the interactive exploration of spatial data from many levels of information granularity, many themes, many epochs and many display modes which are synchronized or not: maps, tables and diagrams.*" (Bédard *et al.*, 2007)

5. 2. b. *State of the Art*

i. Concepts of the Non-Spatial World

OLAP technology stems from the world of multidimensional databases. It was Edgar F. Codd, the father of the relational model, who coined the term and developed the 12 rules governing the OLAP systems (Codd *et al.*, 1993). At the time, Codd was acting directly for Arbore Software, a supplier of multidimensional DBMS. It was therefore not surprising that the first tool on the market was in fact an extension to the multidimensional DBMS from the time, i.e. *Essbase* in 1993. In 1998, *Microsoft* launched his OLAP *Analysis Services* server, which today is the biggest seller.

The research teams of Bédard *et al.* (1997) and Han *et al.* (1998) were among the first to propose a multidimensional approach for the development of spatial data warehouses. The term Spatial OLAP or SOLAP was therefore introduced by Bédard (1997) in reference to the term spatial database. Since the end of the 1990s, several research projects had the objective of combining analytical databases to spatial databases. The pioneers at Simon Fraser University developed the prototype GeoMiner (Stefanovic, 1997), which included an efficient method for the materialization of spatial cubes (Han *et al.*, 1998; Stefanovic *et al.*, 2000). Other researchers at the Université Laval (Bédard, 1997; Rivest *et al.*, 2001) experimented with various combinations of GIS and OLAP technologies in several fields of application (Bédard *et al.*, 2005) before developing the first commercial SOLAP tool: *JMap Spatial OLAP Extension* (Bédard, 2005).

In 2005, OLAP solutions had world-wide sales of 4.9 billion dollars (Source: Optima Publishing). Furthermore, analytical software in general posted a growth in licence sales of 7.3% globally, during the 2004 to 2009 period (source: Gartner)[73].

The use of OLAP tools are perfect for decision needs and make it possible among other things to simplify navigation in the multidimensional database. The OLAP was designed for the rapid and easy exploration of multidimensional data composed of several levels of aggregation (Caron, 1997). Such an approach is very intuitive and makes it possible to construct one's analysis by mainly using the mouse in the manner of hyperlinks (Yougworth, 1995).

OLAP operators make it possible to navigate in the data without having to use a query language. Through simple mouse clicks, it is possible to *drill-down* into the levels of detail and *roll-up* as easily with the drill operators. The OLAP interface makes it possible to view the data in the form of dynamic tables, charts and diagrams, i.e. they directly support OLAP operations. These forms of graphic representations are very flexible and make it possible to easily visualize several dimensions combined in the same view (e.g. cross table).

The general architecture of an OLAP application is composed of three elements: the multidimensional database, an OLAP server and the OLAP client, which allows users to conduct various analyses via a specialized interface and specific operators. It is possible to implement a multidimensional databases based on various approaches: without a server, or with relational OLAP (ROLAP), multidimensional OLAP (MOLAP) and hybrid (HOLAP) server, combining ROLAP and MOLAP (see figure 13).

The OLAP architecture without a server is particular, since it uses *Structured Query Language* (SQL) to access data that is stored in relational tables in the form of a multidimensional view. This original configuration can offer OLAP capabilities, i.e. an adequate response time for queries on aggregates as well as the multidimensional management of data (i.e. dimensions, hierarchies and measures). The principal disadvantage of this architecture relies on the absence of the OLAP server to process the calculation of aggregations. Consequently, all the aggregations required for the application must be pre-calculated and stored in the database. This operation can be very long and produce a significant volume of data. However, the flexibility offered by this architecture, in the absence of the OLAP server, can become necessary in the development of more complex applications or beneficial for certain geomatics applications (e.g. real time), in addition to minimizing the additional technological requirements.

[73] *http://solutions.journaldunet.com/dossiers/pratique/entrepot-donnees.shtml*

The ROLAP architecture also uses relational tools to simulate a multidimensional cube. However, the use of an OLAP server eliminates the main limitation of the architecture without a server. The response time for queries can be longer than the same query conducted using a MOLAP architecture, since the relational indexing methods do not perform as well. Also, the purchase of an OLAP server is an additional expense for the project, which can be a disadvantage in relation to a structure without a server developed on a popular transactional platform (e.g. *Oracle 1og* which uses *Structured Query Language* (SQL)).

MOLAP architecture is the only implementation that results in a physical cube, i.e. a multidimensional database optimized for the storage of multidimensional data. This architecture offers, through the MOLAP server, very powerful aggregation operations, specialized indexing methods and the optimization of the volume of data storage. However, the MOLAP structure and the data handling language (e.g. *Multidimensional Expression* (MDX) for *Microsoft SQL Server*) are sometimes proprietary to the multidimensional tool chosen.

Lastly, HOLAP architecture makes use of two types of structures in order to optimize access to queries on the detailed data stored in the relational structure and the aggregates stored in the multidimensional structure. It should be mentioned that despite the superior performance of the MOLAP architecture, the ROLAP and HOLAP structures, and also the relational structures without OLAP servers, offer exemplary performances that greatly exceed the performances of transactional systems for queries of analytical nature.

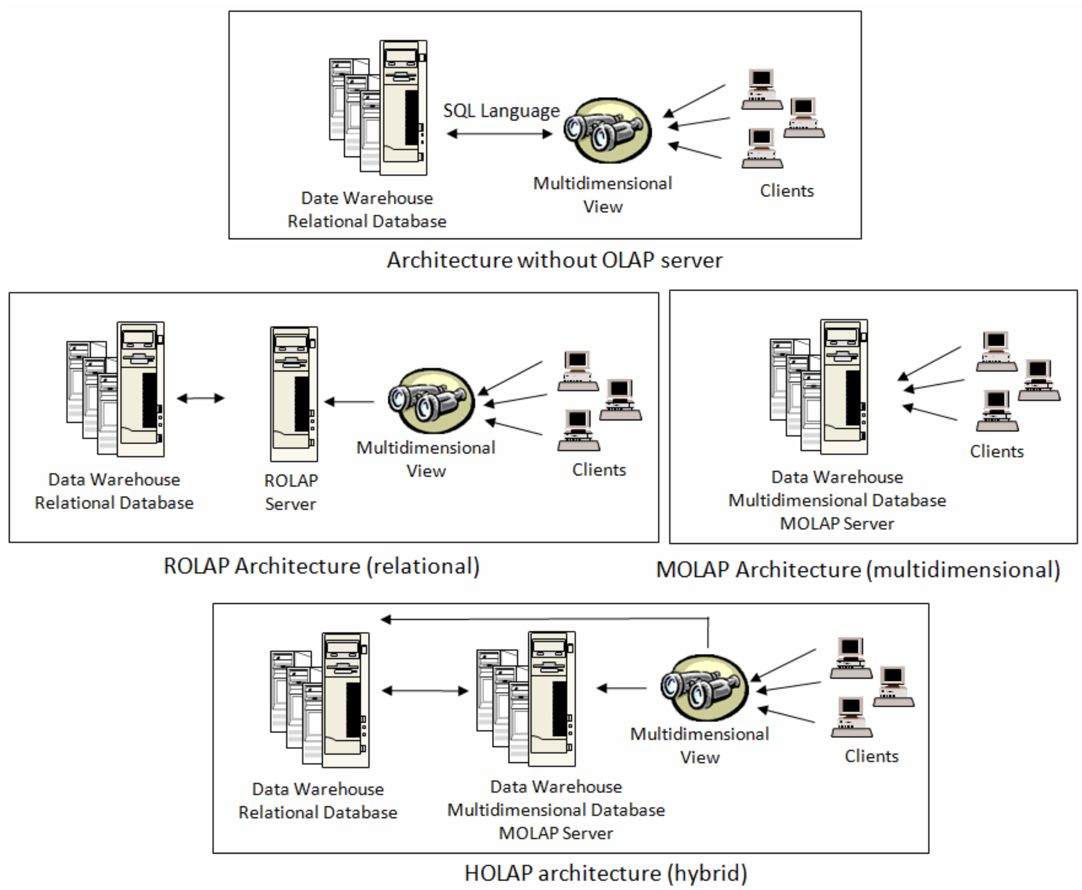


Figure 12. Illustration of the various OLAP server architectures (source: Bédard et al., 1997)

The following table summarizes the various characteristics of OLAP architectures.

Characteristics	Architecture Without OLAP Server	Relational-OLAP Architecture	Multidimensional-OLAP Architecture
OLAP server	Low cost because no OLAP server required	Requires the purchase of a ROLAP server	Requires the purchase of a multidimensional database management system and a MOLAP server
Data structure	Simulates a multidimensional structure	Simulates a multidimensional structure	Supports a physical multidimensional structure
Preparation of aggregate data	Requires a great deal of handling and particular knowledge	The server takes care of generating aggregate data	The server takes care of generating aggregate data
Storage of aggregations	Requires the complete pre-calculation of data	Optimization of data calculated on-the-fly and pre-calculated data	Optimization of data calculated on-the-fly and pre-calculated data
Optimization	None. Requires lots of disk space and pre-calculation time.	Optimizes pre-calculation time, but may require more space than MOLAP	Optimizes disk space and pre-calculation time
Indexing	Uses the transactional indexes	Uses the transactional indexes	Uses higher-performing multidimensional indexes

Table 2. Comparison between the characteristics of relational and multidimensional OLAP architectures (source: Proulx et al., 2004).

ii. Concepts of the Spatial World

On the other hand, Spatial OLAP tools form a new family of interfaces. They are designed like client applications operating data cubes constructed using multi-scale spatial data warehouses or spatial data marts.

The SOLAP tool, in addition to benefiting from OLAP characteristics, allows for the simple and quick displaying of cartographic views. The SOLAP adequately manages the cartographic symbolization rules to be applied on the analysis results. Consequently, the use of such a tool does not necessarily require the support of an expert in cartography, even if it does allow the user to create hundreds of thousands of various maps with a few mouse clicks. In the presentation of the results, the SOLAP technologies use the same rules for graphic semiology (e.g. colour, pattern, border, symbols) for all the displays, whether these are tables, charts or maps. This enables visual synchronization between the various modes of presenting information and to have a homogenous panorama. The graphic semiology used for the various types of displays (i.e. tables, charts and maps) remains synchronized during drilling or during other navigation operations, thereby conserving a perceptual continuity necessary in the discovery of correlations.

SOLAP technology also offers new decision-making assistance functions that are not available in traditional GIS. In addition to allowing the cartographic visualization of data, this technology allows for the interactive exploration of data at various levels of detail, within the map itself, in the superimposed diagrams on maps, in the tables or in the charts, based on various types of drills. The SOLAP tools offer navigation operators for exploring, on the map itself, all the data (e.g. spatial drill-down, spatial drill-up and spatial drill-across). The spatial drill-down operators allow users to navigate from a general level to a more detailed level within a geometric spatial dimension (e.g. view the provinces within a country) and obtain more detailed data when necessary and available. A spatial drill-up operation allows for reverse navigation, i.e. to go up from a detailed level to a more general level (e.g. view the regional values of a grouping of sub-region). Lastly, a spatial drill-across operator makes it possible to view the various members of the same detail level in a spatial dimension (e.g. viewing to better compare the measures from the south region in relation to those from the north). These operators are used directly in the map, in a chart or in a table. The spatial navigation operators can apply to an individual object (e.g. view the administrative regions composing the province (object) of Quebec or apply to all objects of a level of detail (e.g. viewing all the regions composing the provincial level). Several views can be synchronized when needed (e.g. drilling in a map will also drill automatically in the table or pie chart presenting the same data), thereby creating a collection. Several collections can be viewed at the same time and explored independently, in which case the use of a computer supporting two screens is highly recommended.

5. 2. c. ***Positioning in Relation to Other Technologies***

Since OLAP and SOLAP technologies are from the analytical world, it is appropriate to compare them to DBMS and GIS from the transactional world in order to assess the added value.

In the beginning, the technology's capacity to directly query the various levels of data aggregations is a comparison criterion making it possible to distinguish technologies that use the multidimensional approach of others. Then it is necessary to position each tool based on its capabilities to manage the spatial data. Figure 14 illustrates the positioning of the various technologies addressed to date from the point of view of the nature of the information processed and the level of data aggregation.

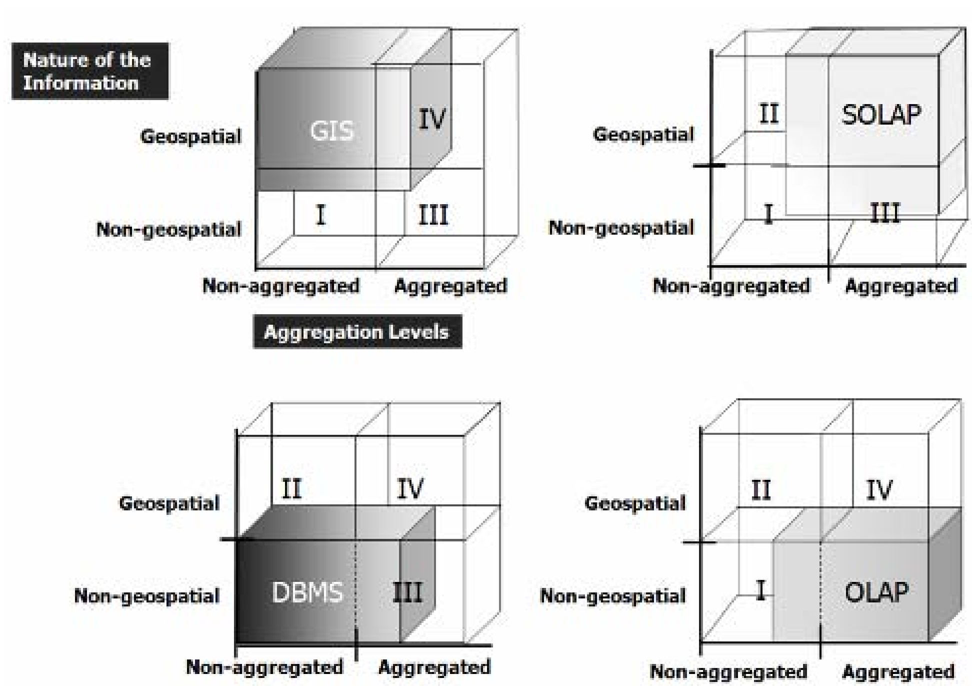


Figure 13. Positioning of the tools based on the nature of the information and its aggregation levels (source: Proulx & Bédard, 2004b).

In reference to the previous figure, in quadrant (I), the DBMS is presented, which enables the management of non-geospatial and non-aggregated data. By extension, it is possible to produce aggregated data with the DBMS since there are summation functions in SQL language, but it is not its principal function. Quadrant (II) positions the GIS, which is used for the management of non-aggregated geospatial data. It is possible to manage non-spatial data in the GIS since it is coupled to a DBMS. It is also possible to generate aggregated data with the GIS using a spatial analysis function such as the merge functionality, but the management of aggregated data is not its primary function. Quadrant (III) presents OLAP, which enables the management of non-geospatial and aggregated data. By extension, it is possible to manage non-aggregated data in OLAP using the drill function towards detailed data (i.e. *drill through* offered by certain ROLAP servers). Quadrant (IV) positions the SOLAP, which enables the management of aggregated geospatial data. Given that this tool is a coupling of GIS and OLAP functions, it is possible to extend the application of the tool to the management of non-geospatial and non-aggregated data.

Furthermore, the SOLAP and the GIS are not mutually exclusive, because the SOLAP is often presented as an additional module to the GIS. The aim of the SOLAP is not to replace the transactional functions of the GIS. The SOLAP facilitates access to detailed and aggregated spatial data by allowing the creation of maps with a few simple mouse clicks, minimizing the use of the keyboard. Spatial OLAP technology supports the multidimensional structure as used in business intelligence, providing it with an immense advantage over web mapping software (e.g. *ArcIMS, GeoMedia WebMap, Push' n see, MapX*), because the latter are based on a transactional data structure.

As presented in figure 15, the GIS and OLAP tools distinguish themselves through their ease of use, the speed in the execution of queries and the number of levels of detail that they natively support. In practice, each has a very specific niche, but the SOLAP tool generally completes the capabilities of a GIS when it is time to conduct analyses.

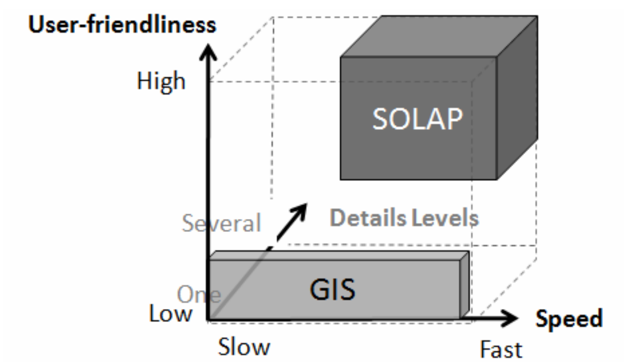


Figure 14. Positioning of the GIS capacities and a SOLAP tool (source: Proulx et al., 2004a).

The following table presents the principle distinctions between a GIS and a SOLAP in regards to the targeted users, the type of data supported, the processes and the architecture.

Characteristics	Geographic Information System (GIS)	Spatial OLAP
Targeted Users		
Type of users	Operational level	Decision-support level
Prerequisite	Knowledge of query language and GIS functions	Be at ease in the hyperlink-type navigation
Learning	Days	Hours
Type of Data		
Finesse of the data	Detailed	Detailed and aggregated
Data temporality	Single time period	Multiple time periods
Process		
Construction of queries	Necessity for heterogeneous handling (see prerequisites)	Homogenous handling by mouse clicks (see prerequisites)

Time of query execution	Variable, based on the complexity of the queries and the number of tables and joins involved	Stable because the query structure is the same regardless of its degree of complexity due to the multidimensional structure
Exploration of data	Discontinued by the use of heterogeneous manipulations and very variable query execution times	Ongoing exploration using drill operators on the data and more stable query execution time
Data Architectures		
Technologies	Relational	Relational or multidimensional
Architecture	Normalized database	Encouraged data redundancy
Data volume	Space minimized by the normalized structure	Large volume generated by aggregation storage
Data source	Comes generally from the acquisition of data	Comes from the integration of several data sources under the form of a data warehouse or data mart
History management	Through data updates	Through the addition of new data (since the history is retained)

Table 3. Principal distinctions between the use of a GIS and an OLAP in regards to the users targeted, the type of data, the processes and architectures supported (source: Proulx & Bédard, 2004b).

5. 2. d. ***Existing Technologies***

i. OLAP Technologies:

In order to help companies and consumers make an enlightened decision regarding all the OLAP products, a study published annually by the OLAP Survey Group, led by analyst Nigel Pendse, presents the market share for Business Intelligence products. The year 2006 marks the 6th edition of the OLAP Survey[74] and this study compiles the Business Intelligence (BI) experiences of 1679 organizations in 87 countries and 32 different categories of organizations. The OLAP Survey is an independent study of the BI field and the results can be obtained through a paid subscription. In 2006, 13 products have a sufficient presence on the world market to be listed in the study. Unfortunately, *SAS OLAP Server* was not retained, given its recent introduction on the market, although it is, in our opinion, among the servers offering advanced capabilities.

A free summary of the study published by Microstrategy (MicroStrategy, 2006) analyzes 6 similar products, i.e.: *BusinessObjects*, *Cognos*, *Hyperion*, *MicroStrategy*, *Oracle Discoverer* and *Business Warehouse* by SAP. The sales of *BusinessObjects* have dropped significantly compared to 2005 and *Cognos* has experienced a slight decrease. *Oracle Discoverer* distinguishes itself by frequently being listed behind the major products since it is more of a report management tool rather than an OLAP tool. The reader is encouraged to consult the publication to gain an overview of the functionalities analyzed in the study.

Table 4 presents the servers and their best-known clients in the North American market. The product's market share is indicated when it appears in the OLAP Survey.

---

[74] *http://www.survey.com/olap/*

Products	Market Share	Servers	Client
Relational OLAP (ROLAP)			
MicroStrategy	7.3%	*Different market servers.*	Microstrategy 8.0
Multidimensional OLAP (MOLAP)			
Hyperion	18.9%	Essbase	Essbase Analytics
IBM	N/A	DB2 Olap Server	*Different market clients.*
SAP	5.8%	Business Warehouse	
Hybrid OLAP (HOLAP)			
Microsoft	31.6%	Analysis Service	Proclarity
Oracle	3.5%	Option OLAP Oracle 10g	Discoverer
Cognos	12.9%	*Different market servers.*	Cognos 8i
BussinessObject	7.3%		BusinessObject XI

Table 4. OLAP Servers and their Client Tools

To these commercial tools have been recently added open OLAP servers and clients. ***Pentaho Analysis Services*** is a client tool interfacing the Relational OLAP server ***Mondrian OLAP Server***[75]. These tools, developed in Java, foresee MOLAP support very soon. However, it is possible to connect (via *Java Database Connectivity* JDBC) to several relational databases. Coupling with *Google Maps* is also possible for map viewing.

Palo-Server[76] is a Multidimensional OLAP specifically developed for the storage of *Enterprise Spreadsheets* data from *Microsoft Excel*, but it also offers connections with the principal databases *SAP, IBM, Microsoft* and *Oracle*.

ii. SOLAP Technologies:

According to the literature, it is possible to classify spatial OLAP tools between three families of solutions based on the available technologies and functionalities, i.e. (1) the OLAP-dominant solutions, (2) the GIS-dominant solutions, and (3) the integrated and hybrid solutions that call upon both OLAP and GIS functions (LGS Group, 2000). At the heart of this classification, the graphical user interface generally comes from the dominant tool and this is the one that offers or that calls upon the minimum functionalities of the other tool. The three solution families respond to various needs. In the first case, the mapping component is only accessory. In the second case, it is the OLAP component that is accessory. In the last case, the two components

[75] *http://mondrian.pentaho.org*

[76]*http://www.jedox.com/en/enterprise-spreadsheet-server/excel-olap-server/palo-server.html*

are deemed important and their coordination or synchronization is a key particularity of this family of solutions.

i. OLAP-dominant coupling:

Despite a decade of research, the products supporting a few OLAP functionalities coupled to a GIS have only appeared on the market very recently. The majority of these products come from key players such as *SAS, ESRI, MapInfo, Business Objects* or *Cognos* (see table 4). This type of solution provides all the functionalities of an OLAP tool. It is therefore implicit that such a solution uses the capabilities of an OLAP server. However, this solution integrates only a few GIS functions, generally the display functions, the map navigation functions (e.g. pan and zoom) and geometric element selection functions (Bédard *et al.*, 2005).

SAS Web OLAP viewer for Java[77] makes it possible to view *SAS OLAP* data on maps produced by *ArcGIS* by ESRI like any other data views. The user can drill into the map's regions to view OLAP information. *SAS OLAP* supports synchronized drilling between the maps and tables.

Cognos 8i[78] makes it possible to view data in TAB format from *MapInfo* structured in the form of a *geoset* with *Cognos Visualizer 1.5* or *MapInfo Location Intelligence*.

BusinessObject XI[79] offers the possibility of integrating a bidirectional bridge between the ESRI geographic data and the BusinessObjects products using a *BusinessObjects Integration Kit for ESRI GIS* or *MapInfo Location Intelligence*.

As for **Microstrategy**, it allows the viewing of spatial data using *MapInfo Location Intelligence*. Other solutions also exist for *Essbase* by Hyperion and *Proclarity* by Microsoft.

ii. GIS-dominant coupling:

The GIS-dominant solutions offer all the functionalities of the GIS tool, but only a subset of the functionalities from the OLAP tool (e.g. limitations in dimension pivot and map drilling). This solution couples a relational database simulating an OLAP server to a GIS software or a spatial data viewing tool. For example, ***OLAP Add-on for ArcGIS (ESRI)*** allows OLAP server users (e.g. *Microsoft SQL Server, SAS OLAP Server* and *SAP BW*) to view their data in the *ArcGIS* environment in the form of a read-only view (i.e. or an excerpt of OLAP data in the form of a

[77] *http://www.sas.com/technologies/bi/query_reporting/webolapviewer/factsheet.pdf*
[78] *http://www.cognos.com/pdfs/factsheets/fs_c8bi_analysis.pdf*
[79] *http://www.france.businessobjects.com/pdf/products/businessobjects_xi_whats_new.pdf*

spreadsheet). The tool offers a utility making it possible to manually connect a saved OLAP view to an ArcGIS mapping layer. The OLAP views must be connected one by one in ArcGIS and navigation is only possible on descriptive OLAP data. Spatial OLAP navigation is not supported. This solution therefore remains the most limited solution.

The simple coupling of GIS and OLAP is not sufficient and many challenges must be overcome to obtain a more efficient solution. Consequently, the development of a commercial spatial OLAP solution makes it possible to integrate all the OLAP and GIS functionalities and to enrich them.

iii. Integrated Solution:

This type of solution, integrating the functionalities of an OLAP and GIS tool, could be qualified as a geospatial-centred application where the geospatial reference of the objects is used constantly in the exploration and analysis of data, in a manner that is as open as with non-spatial dimensions (Bédard *et al.*, 1997). This type of solution is useful when the application must integrate itself in a geomatics environment with a strong data flow (e.g. for data updates, interoperability), requiring the use of functions specific to GIS, such as spatial analysis operators. The solutions from this family are possible using function libraries from OLAP client software and GIS software, or using SOLAP technologies. The integrated solution offers the greatest return on investment (superior flexibility, better user interface, lower costs and much shorter development time).

In the first case, the development of such a solution is possible subject to lots of programming within a specific application framework. In order to do so, certain OLAP technologies such as *ProClarity* from Microsoft and *Essbase* from Hyperion make their function and object libraries available to the specific applications using current programming languages such as Visual Basic or C++. It is therefore possible to develop an OLAP extension to be integrated in a GIS software such as *MapInfo, ArcView* by *ESRI* and *GeoMedia* by *Intergraph*, which enable the use of their function libraries with *MapX, MapObjects* and *GeoMedia* products respectively.

Certain third party tools can also be combined to OLAP servers to produce an interesting integrated solution. This is the case with *MapIntelligence*[80], which offers mapping capabilities from spatial data under *ESRI ArcIMS* and *MapInfo MapXtreme* coupled with data from

[80] *Map Intelligence 2.2 d'Integeo http://www.integeo.com/*

BusinessObjects, Cognos, Hyperion and *Excel*. With *Mondrian OLAP Server*[81], it is possible to implement a multidimensional application using *PostgreSQL*[82], which is an open source relational database management system. *PostGIS* is a spatial extension making it possible to develop GIS applications using *PostgreSQL*. *PostGIS* offers different data types, spatial functions, spatial indexes and a spatial query language. The implementation of *PostGIS* is based on the OGC Simple Features Specification for SQL.

The *MySQL*[83] database management system was initially constructed for transactional applications, a commercial application named *OLAP4ALL*[84] supports multidimensional analyses on *MySQL*. The processing of geographic information is possible using the *MySQL* spatial extension that is based on the OGC Simple Feature Specification for SQL.

We must not forget the open source web mapping technologies discussed in section (4.2.d), which can be an integral part of such a development.

However, a SOLAP technology that is ready for use is a necessity in our field. In this direction, in 2005, a technological transfer between the Université Laval (Dr. Yvan Bédard's research team) and the Quebec company KHEOPS Technologies, owner of the *JMap* solution, allowed for the development of the first SOLAP software available commercially at the international level: ***JMap Spatial OLAP***.

JMAP Spatial OLAP[85] is the first web technology that completely integrates the geospatial dimension in a business-intelligence decision-making environment. It offers an intuitive graphical user interface enabling non-specialists to very easily access their geospatial data in order to view and analyse them. The user interfaces can include several thematic maps, statistical diagrams (bar charts, pie charts, etc.) and tables displayed based on graphical semiology rules defined for classification values or members. The *JMap* mapping server makes it possible to interface the spatial data in a native format and connect to any JDBC-compliant database.

In order to clearly present the integration of a SOLAP technology in the deployment of a corporate solution dedicated to decision-making, figure 16 takes the previous figures and inserts three types of possible couplings for the deployment of a SOLAP technology as an analytical client.

---

[81] *Mondrian official home page, http://mondrian.sourceforge.net*
[82] *Postgresql official home page, http://www.postgresql.org*
[83] *Mysql official home page, http://www.mysql.com*
[84] *OLAP4all, http://www.olap4all.com/*
[85] *JMAP SOLAP, http://www.kheops-tech.com/en/jmap/solap.jsp*

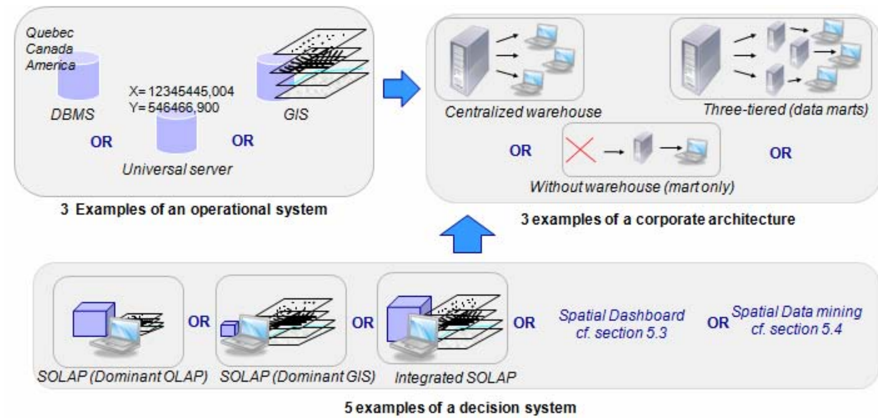


Figure 15. Positioning of possible SOLAP technology couplings in the deployment of a corporate architecture dedicated to decision-making.

5. 2. e. ***Examples of Health Applications***

Apart from an OLAP application using *Oracle Express* developed using the data warehouse of the Public Health Institute of the Republic of Slovenia (Hristovski *et al.*, 2000), nothing else was identified in our search of the literature as an OLAP application in health.

It is new that spatial OLAP applications are implemented in organizations for the daily management of decision-making. Various organizations have developed SOLAP applications in Canada, France, Portugal, Brazil, the US, Italy, etc., but very few have done so using mature SOLAP technology, since these installations were prototypes or rather limited *ad hoc* developments.

Although the study undertaken for Health Canada in 2000 (Gosselin *et al.*, 2000) presents several pilot projects of the *National Health Surveillance Infrastructure (NHSI)* that combined GIS and OLAP (e.g. SPHINX, GISII), we were unable to find any trace of them six years later in the scientific literature and the Internet.

Though there is a need for aggregated health-related spatial data, few applications are in operation outside of those developed within research institutions, i.e. ICEM/SE (Proulx *et al.*, 2002; Bédard *et al.*, 2003), SOVAT (Scotch & Parmanto, 2005) and GOLAPA (da Sylva *et al.*, 2006). In the three previous projects, an environmental health prototype was put in place as a proof of concepts.

The project ICEM/SE (Interface Cartographique pour l'Exploration Multidimensionnelle des indicateurs de Santé Environnementale sur le World Wide Web [Mapping Interface for the Multidimensional Exploration of Environmental Health Indicators on the World Wide Web])

(Proulx *et al.*, 2002; Bédard *et al.,* 2003) made it possible to improve decision-making and interventions in the field of environmental health by facilitating the development and validation of new research hypotheses. ICEM/SE has led to the development of a spatial data analysis tool by facilitating the temporal and multi-scale management of data through the use of the OLAP approach. Since then, the ICEM/SE prototype has given way to the first declared invention of the Canadian GEOIDE[86] Network. The technological transfer towards the industry was completed with the firm KHEOPS Technologies for the marketing of a commercial product (see section 5.2.d.ii existing technologies). For its part, SOVAT integrates OLAP (based on *Microsoft SQL Serveur 2000* and *VB.NET*) and GIS functionalities to facilitate the geographic knowledge discovery process in a user interface that is easy to use for health surveillance needs. For its part, GOLAPA (*Geographic On-line Analytical Processing Architecture*) is based on the development of an open and extensible architecture for the geospatial and multidimensional processing of data, based on the open source Mondrian OLAP server.

In Canada, *Cancer Care Ontario* has, since June 2006, deployed the second implementation of a data warehouse on the web and the ***iPort***[87] analytical tool. Developed using *MicroStrategy 8*, the mapping aspect was assured by *MapInfo Intelligent Mapping* based on the documentation (TDWI, 2006).

In Quebec, in 2004, the Public Health Directorate developed the Plan intégré des resources informationnelles en santé publique (PIRISP) integrating the contribution of all players in the health system and social services, including the Infocentre de santé publique (Government of Quebec – ministère de la Santé et des Services sociaux, 2004). The PIRISP identified the Infocentre de santé publique as a priority for the dissemination of data from the common surveillance plan in the form of an Intranet. The project supervisor, the Institut de santé publique du Québec (INSPQ), is currently studying SOLAP technologies in order to assess their usefulness and their positioning in relation to other Infocentre tools for the analysis of air and water quality data in Quebec. The INSPQ is also supporting other SOLAP applications contributing to improving environmental health surveillance by providing access to twenty or so health and environmental indicators (INSPQ, 2006). The OLAP-Contamination application, which allows for the use of data on bivalves, echinoids, birds, fish and mammals contaminated with metals and organochlorines in Quebec's North is an example. Also, a joint research project between the Centre for Research in Geomatics of Université Laval, Health Canada, the INSPQ and Ouranos, funded as part of the GEOIDE program, proposes a first interactive web tool for better understanding the health vulnerabilities related to climate change (Badard *et al.*, 2006). The

---

[86] *Réseau GEOIDE, http://www.geoide.ulaval.ca*
[87] *http://cancercare.on.ca/index_requestDatafromCCO.htm#how_do_i_request_data*

project focuses on the development of a data warehouse, web services and a SOLAP application enabling the exploration, summary and online analysis of spatio-temporal data linked to climate change, in order to better understand health impacts and vulnerabilities.

Lastly, the Groupe de recherche interdisciplinaire en géomatique de la santé PRIMUS at the Université de Sherbrooke is currently working on a Système d'Information Spatio-Temporel sur les maladies chroniques [Spatio-temporal Information System on Chronic Diseases] and first-line care (SIST-MC). The objective of this project is the integration of advanced functions for spatial and temporal analyses adapted to the study of chronic diseases in primary care. A Spatial OLAP application was deployed as proof of concepts in order to evaluate if such a tool would be of interest to potential users such as government organizations, regional authorities, foundations. The prototype was presented at the 9th GEOIDE Annual Scientific Conference in Halifax (Vanasse *et al.*, 2007).

5.3 SPATIAL DASHBOARD

5. 3. a. *Summary Definitions*

"*A **management dashboard** is a means of selecting, putting together and representing essential and pertinent indicators, in a summarized and targeted manner [...] providing both a global vision and the possibility of mining in the levels of detail.*" (Voyer, 2000; p.39).

"*A **dashboard** is a performance measuring instrument enabling the "pro-active" piloting of one or more activities as part of a progress process. The dashboard contributes in reducing uncertainty and facilitates the inherent risk taking in all decisions. The dashboard is a decision-making assistance tool.*" (Fernandez, 2007).

In the field of decision-support systems, the executive dashboards are also referred to as scorecards, balanced scorecards, scoreboards, steering panels, control panels or executive information systems.

Curiously, the *dashboard* definitions do not mention any technological content. This suggests that the dashboards are generally disseminated in printed form. The definition of dashboard at the Office de la langue française (OLF, 2007) speaks of a summarizing document. However, the definition of digital dashboard extends its application to a "more developed information management tool destined to office computers." However, the definition that approximates most what is meant here by dashboard is the *executive information system*.

"The ***executive information system*** *provides summarized and up-to-date information that provides an ongoing general overview of the company's activities and operations, based on external and internal sources. It is a computerized dashboard for senior executives that is used in strategic planning and based on which reports, charts, etc. can be generated and can be consulted quickly.*" (OLF, 2007)

"*A **spatial dashboard** is a tool that presents a few simple, explicit and structured views, including maps, of the key indicators (spatial and non-spatial) useful for the user. It also gives access to detailed views that allow a better understanding of phenomena and to prospective views that allow for predicting what could happen in the future.*" (Bédard *et al.*, 2007)

5. 3. b. *State of the Art*

In the beginning, the dashboard refers to automotive equipment that allows drivers to know their vehicle's status in terms of speed, oil level and fuel. Historically, the idea of dashboards developed itself in the work done in the 1970s on decision-making assistance tools. In the 1980s appeared a computerized version of the dashboards, the executive information systems or EIS). However, these dashboards were not very successful since they were developed on central computers that were costly to maintain and deploy (Eckerson, 2005).

At the end of the 1990s, with the non-stop development of the web, the need for information dissemination began to appear, at the same time as new disciplines such as balanced scorecards and six sigma took on more importance in estimating business performances. At that time, digital dashboards as we know them today began to appear. In the beginning, many systems were constructed by the piece to respond specifically to the demand of organizations. Today, digital dashboard technology is available via integrated products offered by several software manufacturers. However, certain companies continue to internally develop and maintain dashboards, by offering consultation services in addition to the software solution. Of course, the dashboard, like any decision-making assistance tool is dependent on the proper analysis of needs and only experienced companies can properly respond to this necessity.

The dashboard is a decision-making assistance tool that is quite popular. In 2004, a study by The Data Warehousing Institute (TDWI) revealed that 50% of organizations already use a dashboard and 17% are currently in the process of developing one. The same study stated that one third of organizations that possess a dashboard used it as a principal application for data analysis.

Today's dashboard is computerized; it presents a few simple, explicit and structured views of the user's key indicators. It is generally composed of several performance indicators illustrated through more generalized views (icons, dials, gauges) or through more detailed views (histograms, pie charts, tables and more recently maps). The dashboard is also characterized by the various themes that can be combined (a bit like OLAPs) to generate the displays.

The dashboards make it possible to follow and anticipate the operating of the phenomenon being analyzed. The indicators generally present the progression of a phenomenon based on a possible target or a set objective. A dashboard also includes detailed views to better understand the situation and prospective views to anticipate what might happen (Fernandez, 2005). Drilling tools, unique to OLAP technology, make it possible to navigate in the views and make it an interactive tool much more than a summarizing document, without offering all the flexibility and power of an OLAP tool (which is also much more complex to master).

5. 3. c. ***Positioning in Relation to Other Technologies***

The dashboard is an analytical application that draws from GIS technology or web mapping in order to present decision-support data on a mapping media and use the spatial analysis functionalities of these tools to manipulate the data. Over the last few years, several dashboards have been offering mapping components. However, in order to properly present the characteristics of a dashboard, it is preferable to compare them to decision-support type tools. This way, the comparison with the OLAP tool is more appropriate.

The first distinctive element of the two applications is the data updates frequency. The dashboard requires current data, most often used in real time, in order to allow for rapid decision-making. The OLAP application, for its part, can use data generated with a delay of a few hours, a few days or a few months. In both cases, it is difficult for the displays that make use of data cubes to be in real time, but given that a dashboard also opens the door to the querying of transactional data (e.g. stock exchange lists disseminated over the Internet), whereas OLAP only uses data cubes, the dashboard offers more versatility in obtaining sensitive information in real time.

Also, a distinctive element of dashboards is that, contrary to an OLAP application where the application typically allows access to multidimensional cubes only, the dashboard allows access to an array of data that is both transactional and multidimensional (for the price of analytical power found in OLAP tools). The dashboard can also provide direct unique access to data from various systems within the organization (whereas the access is indirect for data from multidimensional cubes, both for the dashboard and OLAP).

In an OLAP application, all of the business-intelligence data is accessible based on the profile (security rating) of the user and several views, with the assistance of a certain number of data exploration operators. However, with the dashboard, we focus on a few data views considered to be the most significant for the business-intelligence process as well as on a reduced number of operators. Access to simple and explicit views of business-intelligence indicators makes the decision process easier. This simplicity of the dashboard, both through its content and the available functions, makes it possible to better understand a phenomenon as a whole based on a few indicators. This also simplifies the user interface, which becomes better adapted to decision-makers. However, the choice of the right indicators is a critical element for the dashboards, because, by definition, they do not present an exhaustive picture of the situation. Moreover, the voluntarily reduced amount and the typical unavailability of detailed data limit the scope of the analyses using the dashboard.

In an OLAP application, data navigation is executed based on the user's way of thinking in an open and exploratory approach (e.g. in the search for correlations, trends and exceptional cases). In contrast, dashboard navigation is more structured and defined by the way of thinking of the organization or the dashboard builder. The dashboard views are organized based on a thought-out analysis sequence for the user, more predefined than for OLAP. This analysis sequence pursues a precise goal and is optimized to facilitate the securing of responses to a specific question. Inversely, the navigation sequences in OLAP are open to the user and can be used for a larger number of queries. In other words, the connections between the panoramas of a dashboard are predefined to reflect the culture of the company in its strategic analyses, which makes it possible to simplify the user interface and to create a normative effect in the organization. However, the dashboard loses the flexibility of navigation offered by the OLAP tool.

Lastly, it is possible to develop a hybrid solution, combining the two types of decision-support applications where the dashboard is the principal user interface and where a window provides access to an OLAP tool to offer all the desired flexibility for data exploration.

Therefore, these two types of decision-support applications are not geared to the same clientele of users. In looking at the decision pyramid, it can be concluded that the dashboard is aimed typically at strategic-level users (executives and managers) and that OLAP is aimed typically at a tactical level (analysts). Whereas the dashboard is more structured, it is aimed at executives and decision-makers for whom the analyses are repetitive and standardized within the company. In addition, computer knowledge or the interest in technologies being more limited to a strategic level, this favours the framework provided by a dashboard. Inversely, the users targeted by the OLAP application are more often specialists who have various analytical needs, that are difficult to program and often requiring the open exploration of data. For them, the flexibility offered by the OLAP application is indispensable. Table 5 succinctly presents the differences between the dashboards and the OLAP tools.

Characteristics	OLAP	Dashboard
Frequency of data updating	Delayed by a few hours or months depending on the application	Real time, immediate access to data
Data views	Complete views of the data: Several indicators Aggregated and detailed views of the data	Partial views of the data is significant: Reduced number of indicators No detailed views
Types of analyses	Various analyses, non-programmable and often requiring the open exploration of data	Repetitive and standardized within the company
Navigation	Open and exploratory navigation based on the user's way of thinking	Structured analysis sequence defined based on the organization's way of thinking
Type of data	Multidimensional data	Transactional and multidimensional data
Level of integration	Transactional technologies (reporting, GIS, web mapping)	Transactional technologies (reporting, GIS, web mapping) decision-support technologies (OLAP, data mining).
Clientele of users	Tactical level: specialists in the field	Strategic level: senior managers

Table 5. Positioning of dashboards in relation to OLAP tools

The spatial dashboard can also be part of the deployment for a corporate solution dedicated to decision-making. Figure 17 takes the architecture from the previous figures and inserts the spatial dashboard as an example of an analytical client. It can interface components that are both analytical (e.g. OLAP warehouse and cube) and transactional (GIS and DBMS) via the same client.

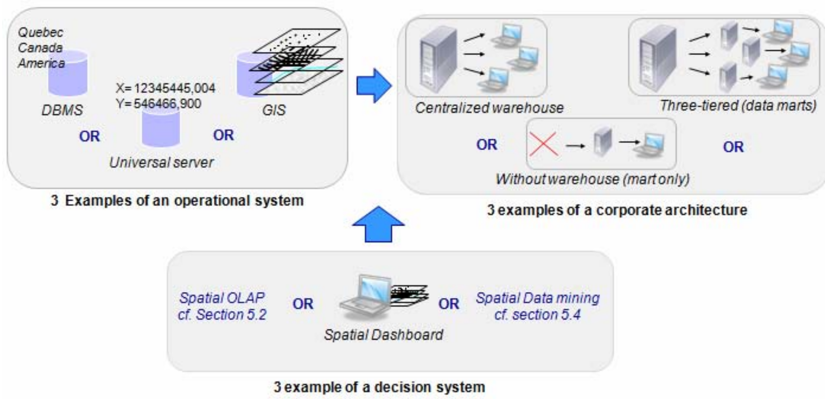


Figure 16. Positioning of the spatial dashboard in the deployment of a corporate architecture dedicated to decision-making.

5. 3. d. ***Existing Technologies***

Various technologies can be used for the production of management dashboards. According to Voyer (2000), there are at least 3 categories, such as spreadsheets, report generating tools and dedicated technologies.

First of all, the various market spreadsheets (e.g. Excel by Microsoft) can make it possible to easily generate dashboards. Various works[88],[89] and Internet sites [90,91] present the concept of a dynamic pivot table, which makes it possible to generate a summary of a raw data table. Various third-party tools propose interesting tools drawing from this approach. Even if Thomsen (2002) suggests that spreadsheets cannot adequately provide all the OLAP functionalities, it is a technology that is easy to implement and can rapidly meet simple needs (Artnick *et al.*, 2002). Artnick *et al.* (2002) even experimented with the technology on health data in Slovenia.

Secondly, the specific reporting modules for a database management system make it possible to directly interface with the database of a third party or the same series of products, e.g. *Office Performance Point Server 2007* by Microsoft, *Crystal Application* by Crystal Decisions, *Esperant*, *Media* by Speedware and *Pentaho Dashboard* by Pentaho. Certain products offer mapping capacities such as *Business Intelligence 8.0* by Cognos, which integrates the web mapping tools by ESRI ArcIMS to the *Metrics Manager* dashboard tool.

Lastly, the dedicated systems possess advanced functions for data extraction and consolidation, break down, summation, and the summary presentation of tables and charts.

---

[88] *Dashboard Reporting With Excel (C.W. Kyd) http://exceluser.com/software/landoffice.htm*
[89] *L'essentiel du tableau de bord [Dashboard Essentials](Fernandez, 2005)*
[90] *Pivot tables, http://peltiertech.com/Excel/Pivots/pivotstart.htm*
[91] *Pivot Table, http://lacher.com/toc/tutpiv.htm*

These tools are often part of a series of BI products that are coupled with both transactional and multidimensional (OLAP) approaches. There are different products such as:

- *Proclarity Analytics 6.0 Dashboard Server* by Microsoft;
- *EIS* by SAS ;
- *Oracle Balanced Scorecard, Oracle Bi Discoverer* by Oracle;
- *Hyperion Performance Scorecard* by Hyperion;
- *Syntell 4i* [92] by Syntell, which can use JMap as a mapping server;

In the Voyer categories, we can add the third-party tools which have carved out a niche in the market and which are in fact tools that can be coupled with market DBMS or OLAP servers, e.g. *Dundas dashboard*[93]. *Dundas dashboard* offers *Map, Chart, Gauge, Diagram* and *Calendar* tools in order to develop dashboards using Microsoft Excel, Microsoft Reporting Services or .NET and ShapeFile spatial data (ESRI). The GIS functionalities in the application are limited to zooming in and out and panning. The application was developed using AJAX (Asynchronous JavaScript and XML). Dundas also supports an OLAP service for the creation of OLAP views in the dashboard.

5. 3. e. ***Examples of Applications in Health***

At the regional level, the Carrefour montréalais d'information sociosanitaire has developed a ***strategic dashboard***[94]. This web tool presents the strategic indicators selected for following the evolution of priority transformation in the Montreal health and social services network. This interactive dashboard allows for the successive display of three viewing modes (radar, table and chart) and the associated documentation makes it possible to properly understand each indicator (definition, use, warnings, methods, measuring units, etc.). This dashboard does not present any map displays; however, the ***Atlas Santé Montréal***[95] by the same supplier (see page 26) helps complete the information using web mapping. However, it is difficult to know, without any additional documentation, if this dashboard uses a multidimensional database. Figure 18 illustrates a radar diagram of the accommodation requests for people 75 and older in three CSSS.

---

[92] *Syntell 4i, http: //www.syntell.com/*
[93] *Dundas Dashboards http://www.dundas.com/*
[94] *Tableau de bord stratégique http://www.cmis.mtl.rtss.qc.ca/fr/performance/tableaudebord/tb_presentation.html*
[95] *http://www.cmis.mtl.rtss.qc.ca/fr/atlas/atlas_presentation.html*

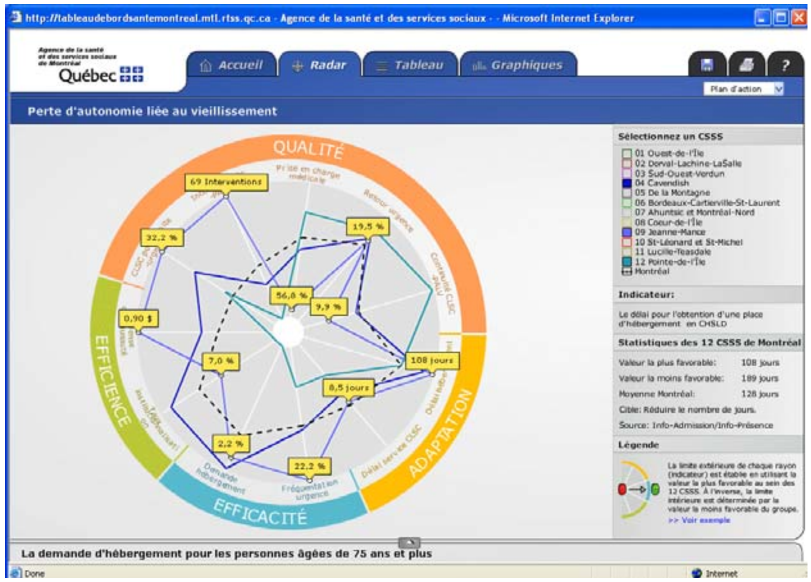


Figure 17. Strategic dashboard by the Carrefour montréalais d'information sociosanitaire presenting the requests for accommodations by people 75 and older in three CSSS.

Few dashboards were identified in the scientific literature review and our web searches. Practically no spatial dashboard was found in the health field. Consequently, to adequately illustrate a spatial dashboard, figure 19 presents an example of a map production mapping application developed by the Université Laval and Syntell for the ministère des Ressources naturelles et de la Faune du Québec (MRNFQ).

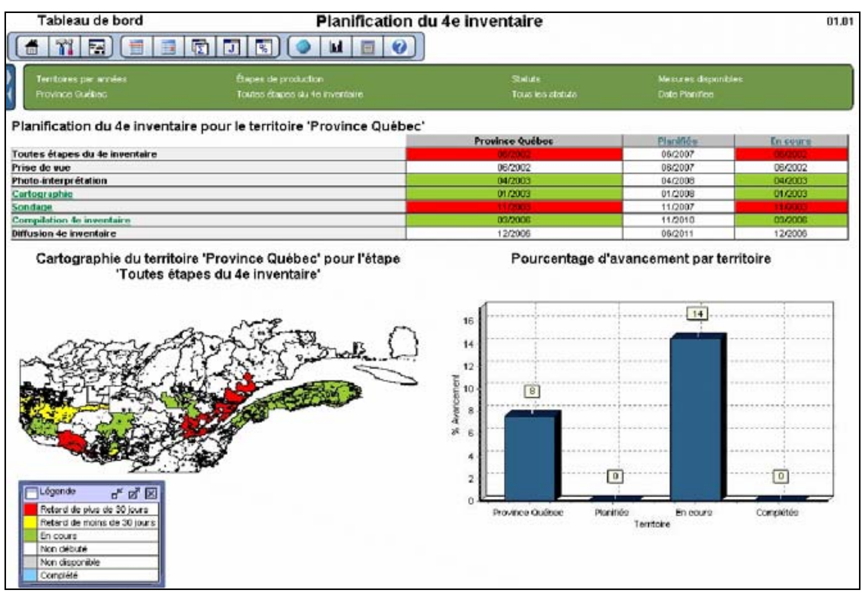


Figure 18. Application for map production management developed by the Université Laval and Syntell for the MRNFQ.

This spatial dashboard is composed of a synchronized OLAP table, map, and histogram. Certain drill operators are provided to facilitate access to the detailed data of the application. The application uses a *Microsoft SQL Server* multidimensional database and was developed using *Syntell 4i* technology.

5.4 SPATIAL DATA MINING

5. 4. a. *Summary Definitions*

Data mining *is a valid and usable knowledge extraction process based on large data volumes.*

"Data mining is a non-trivial process that consists in identifying, in data, diagrams that are new, valid, potentially useful and especially understandable and useable." (Fayyad *et al.*, 1998)

"Category of tools making it possible to automatically extract interesting and intelligible knowledge from databases (rules, regularities, patterns, etc.) and to discover implicit models." (Fayyad *et al.*, 1996).

Spatial data mining *is the "extraction of implicit knowledge, spatial reports and other models that are not explicitly stored in geographic databases."* (Miller & Han, 2001).

5. 4. b. *State of the Art*

The advent of transactional systems, the ever growing capacity of computer systems matched with a greater efficiency in the transmission networks make it so that organizations today have accumulated an impressive quantity of data. It is such a large quantity that it exceeds human analytical capacities. Automatic processes are therefore necessary to extract the useful information (i.e. knowledge) from this raw data. Within this optic, data mining is used to automatically extract interesting knowledge buried in large-volume databases (Gardarin, 1999). Such tools are used today in several fields, ranging from finance (e.g. market analysis and forecasting, fraud detection) to marketing (e.g. analysis and forecasting of the consumer behaviour, prediction of responses to direct marketing operations), bioinformatics, web site development and climate change.

Based on the vision of several authors, data mining is part of a global knowledge discovery in databases (KDD) process (Fayyad *et al.*, 1998). Figure 20 presents this process.

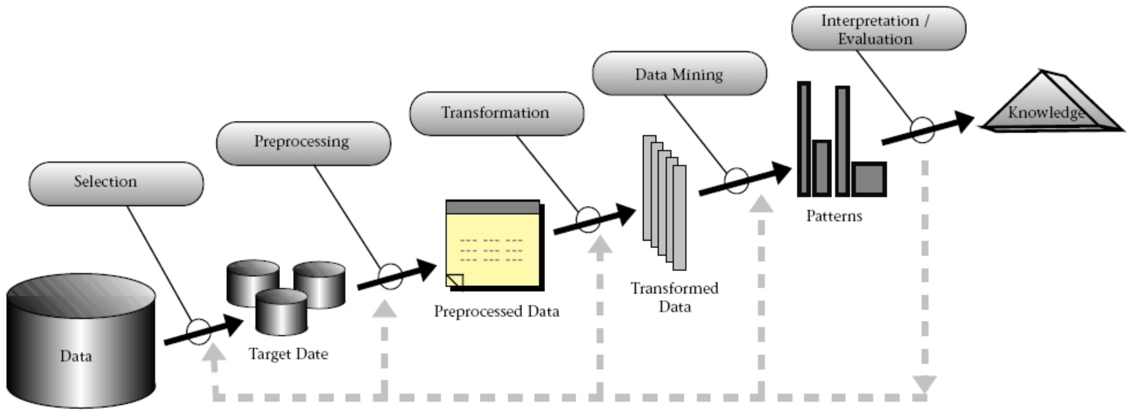


Figure 19. Data mining is part of the global knowledge discovery in database (KDD) (Han & Kamber, 2001)

According to Han & Kamber (2001), data mining can be seen as the natural evolution of information technologies, more precisely the field of databases. Thaerling (Thearling, 2007) proposes the following table to illustrate this evolution.

Evolution	Ex. of Question	Technologies	Companies	Characteristics
Data acquisition (1960s)	"What was my total revenue in the last five years?"	Computers, disks and magnetic tapes	IBM, CDC	Retrospective, static data delivery
Access to data and database management systems (1980s)	"What were unit sales in New England last March?"	Relational databases (RDBMS), Structured Query Language (SQL), ODBC	Oracle, Sybase, Informix, IBM, Microsoft	Retrospective, dynamic data delivery at record level
Data warehouse & decision-making support (1990s)	"What were unit sales in New England last March? Drill down to Boston."	On-line analytical processing (OLAP), multidimensional databases, data warehouses	Pilot, Comshare, Arbor, Cognos, Microstrategy	Retrospective, dynamic data delivery at multiple levels
Data mining (emerging)	"What's likely to happen to Boston unit sales next month? Why?"	Advanced algorithms, multiprocessor computers, massive databases	Pilot, Lockheed, IBM, SGI, numerous startups (nascent industry)	Prospective, proactive information delivery

Table 6. Evolution in the field of databases (Thearling, 2007).

Generally, there are two major categories of data mining: descriptive and predictive (Han & Kamber, 2001). The descriptive processes are used to characterize the general data properties. They make it possible, among other things, to discover unexpected patterns and unsuspected correlations such as, for example, the fact that non-connected products are often bought at the same time. Commercial strategies can then be defined in order to maximize this established fact. The predictive processes produce predictions based on existing data. They are used to predict trends or behaviours and are largely used in the marketing field in order to maximize their return on investment. For example, data mining tools can be used for advertising mail-outs in order to determine the population segment most susceptible of responding favourably to the mail-out.

These various processes are based on the use of several functionalities offered by data mining tools, including:

- The description of the classes or concepts that make it possible to describe the data in a summarized, precise and concise manner. The results of this task can notably take the form of charts, multidimensional tables and discriminating rules.
- The classification and prediction that are used to define a set of properties that characterize and distinguish each class, so as to predict the class of a new value. The results can be analyzed in the form of decision trees and neural networks.
- The grouping that makes it possible to identify the homogenous data groups.
- The association that aims to extract correlations between data.

Each functionality is accomplished via the application of various algorithms. Today, there is a large quantity of algorithms and each is defined based on a specific objective. Unfortunately, there is no universal algorithm that responds to all of the needs and it is often necessary to combine them in order to obtain interesting results and performances.

5. 4. c. ***Spatial Data Mining***

Contrary to the techniques mentioned previously, spatial data mining techniques take into account the nature and spatial specificities of data. Notably, they support the discovery of spatial relations and the optimization of spatial queries (Han & Kamber, 2001). Spatial data mining techniques are relatively more complex to implement than their non-spatial pendants, given the inherent complexity of spatial data. In fact, they must consider the fact that the data are interconnected (contrary to non-spatial data mining techniques that presuppose independent data), that the spatial relations are often implicit (i.e. not stored in the database) and numerous (topological, metric). This results in several challenges and creates a separate research field that integrates concepts stemming from spatial statistical knowledge, spatial databases and traditional data mining as presented in figure 21.

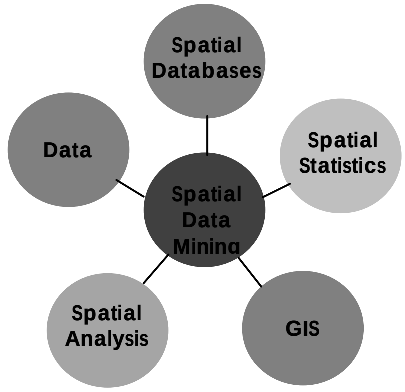


Figure 20. The principal fields that are the basis of spatial data mining

Spatial data mining can be summarized into two principal phases according to Aufaure *et al.* (2000): the exploratory phase, which makes it possible to explore the data to describe a spatial phenomenon and the decision-making phase which makes it possible to explain and predict localized phenomena by looking for correspondences with the properties of the geographic environment.

Lastly, data mining tools, spatial or not, are useless without good knowledge and understanding of the data and do not replace the knowledge of decision-makers. This knowledge is notably necessary in order to distinguish the patterns that are useful from those with no concrete meaning. This is particularly true in the spatial field where geographic constraints (e.g. topography, networks) are not necessarily taken into consideration (Bédard *et al.*, 2007).

5. 4. d. ***Positioning in Relation to Other Technologies***

Compared to spatial analysis techniques typically used in the GIS field, spatial data mining techniques are conducted in an automatic manner, on voluminous databases and within an exploratory context. Table 7 directly compares the spatial analysis with spatial data mining.

Spatial Analysis	Spatial Data Mining
Visual discovery of knowledge	Automatic discovery of knowledge
Confirmatory	Exploratory (generates hypotheses)
Inapplicable in voluminous DB	Operates on large volumes of data

Table 7. Comparison of the spatial analysis and spatial data mining (Zeitouni, 2006)

Data mining tools are not systematically integrated in a data warehouse architecture. In fact, the data sources used in these tools can be of different natures (e.g. text files, web pages, Excel files, relational databases). However, the data warehouses are ideal sources of data mining operations by integrating, within a single repository, data that are cleaned, historic, detailed and aggregated. According to Inmon (1996b), "*data mining can be done where there is no data warehouse, but the data warehouse greatly improves the chances of success in data mining*."

The major distinction between data mining tools and OLAP tools is at the process automation level. The mining processes are based on algorithms and are conducted in large part in an automatic manner whereas the OLAP processes are guided by human intervention and are conducted in an interactive manner. In general, the OLAP tools make it possible to confirm intuitions, whereas data mining tools are used to search for unexpected and non-evident correlations. They are necessary when manual exploration, as achieved with OLAP tools, is practically impossible.

Spatial data mining is also part of the deployment of a corporate solution dedicated to decision-making. Figure 22 takes the architecture from the previous figures and inserts spatial data mining as an example of an analytical client. Spatial data mining can be combined with systems that are both analytical (e.g. warehouse and OLAP cube) and transactional (e.g. GIS and DBMS).

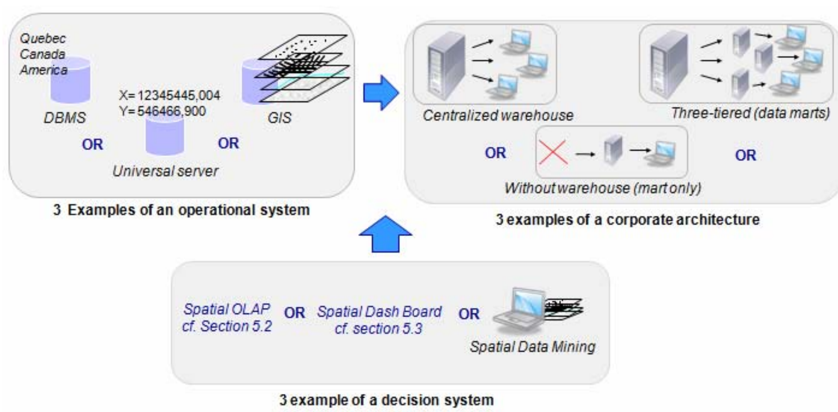


Figure 21. Positioning of spatial data mining in the deployment of a corporate architecture dedicated to decision-making.

5. 4. e. ***Existing Technologies***

Several data mining solutions currently exist on the market. Among the most popular are the solutions offered by *SAS* (Enterprise Miner[96]), *SPSS* (Clementine[97]), *Oracle* (Data miner[98]), *Insightful* (Insightful Miner[99]) and *IBM* (DB2 Intelligent Miner[100]). An open-source solution, Weka[101] by Pentaho is also available.

---

[96] *http://www.sas.com/technologies/analytics/datamining/miner/*
[97] *http://www.spss.com/clementine/*
[98] *http://www.oracle.com/technology/products/bi/odm/odminer.html*
[99] *http://www.insightful.com/products/iminer/default.asp*
[100] *http://www-306.ibm.com/software/data/iminer/*
[101] *http://sourceforge.net/projects/weka*

From a spatial point of view, although several spatial data search algorithms have been proposed, there are no commercial tools to our knowledge. The proposed algorithms are implemented within prototypes. For example, the prototype GeoMiner (Han *et al.*, 1997), an extension of DBMiner (Han *et al*., 1996) uses the capabilities of the *MapInfo* technology in order to allow for searches in spatial data. It takes advantage of a multidimensional approach and the search operations are conducted via an extended-SQL engine called GMQL (GeoMiner Query Language).

In the health field, one of the research teams at the *Commonwealth Scientific and Industrial Research Organisation (CSIRO)*[102] is developing data mining tools specifically adapted to health data. Their product, the *Data Mining Toolbox*, facilitates the execution of the main data mining tasks, notably those preceding analysis, in order to automatically identify the particular problems with the data so they can be cleaned.

---

[102] *http://research.cmis.csiro.au/edm/health.html*

6. FUTURE RESEARCH AND DEVELOPMENT PERSPECTIVES

The democratization of technology in the geomatics industry is leading us gradually towards the development of more user-friendly and accessible technologies. Mass mapping tools on the Internet, such as Google Earth and Google Maps, are progressively becoming popular tools for creating travel itineraries and looking for addresses. Technologies based on the Global Positioning System (GPS) such as vehicle navigation tools and fleet monitoring systems place very specialized technologies in the hands of non-experts.

However, geospatial decision-support systems still pose several challenges and are the subject of a great deal of university research. According to Bédard *et al.* (2007), "*The penetration of BI into geoinformatics started in the mid-90s in a few university research centers. Nowadays, several researchers and practitioners have become active in analytical geoinformatics. Theoretical concepts have become a corpus of their own. System development methods are being adapted and the quality of aggregated spatial data has become a research topic. Spatial ETL is being tackled within universities although it remains the major challenge. These indicators show that the merging of BI and GIS is maturing but that further research will facilitate the flow of spatial data from the geospatial data sources to the Spatial BI tools, and will facilitate the evaluation of the quality of the results.*"

Therefore, research focusing on geospatial databases can be articulated around various aspects including the **design and creation of geospatial data cubes** and the **evaluation of their quality.** Research is in progress in order to improve the formalisms making it possible to produce conceptual models of more sophisticated multidimensional data cubes, supporting among other things various types of spatial dimensions, states and spatio-temporal evolutions, 3D, real time analysis and mobile applications. When compared to the traditional GIS research, research in spatial data warehouses and spatial data cubes is faced with particular problems such as the efficient integration of time management (which is indispensable to data cubes) with space (very few spatial databases are temporal). Management of the evolution of administrative limits (e.g. municipal mergers, changes in the enumeration areas at Statistics Canada) is a good example of the temporal considerations that health professionals must face.

Until now, only the vector structure has been studied to support the spatial dimension in the data cubes. It is therefore a good idea to explore the use of spatial dimensions based on a raster structure[103] to improve the integration of data that is greatly heterogeneous from a temporal, spatial and semantic point of view. The potential of analysis functionalities specific to raster data structures: costs matrix, filtering, analysis of optimal zones, map algebra, as a means of aggregating spatial data, is also being studied.

Several technological advances over the next few years will be based on the development of solutions that are accessible "everywhere", "for everyone" and "at any time". It will then be possible to provide decision-making information at the proper level of aggregation based on the user location, and inversely, to allow the system to automatically send information to the right level of analysis based on the user's movements. These technologies will make it possible to go beyond transactional location-based services, taking into account hundreds of users in transit (on-the-fly aggregation of information) and hundreds of locations travelled by these users (e.g. send an alert to the coordinator of an operation when more than 100 rescuers are in the danger zone, allowing the coordinator to drill down on these zones to decide on the number of rescuers to evacuate).

Location-based services include several research issues in terms of the nature of the decision to support, the functions to offer, the data architecture rules and optimal software, bandwidths, number of tiers, types of geospatial dimensions and lastly the alert modes possible for the user in relation to the data status (e.g. quality, update). Several research projects are currently underway to provide new solutions both in the wireless technologies research field and in geomatics.

In addition to being able to connect to the system from any area, the executives want to have access to the most current data possible. In order to allow the deployment of spatial **data cubes in real time,** it is necessary to develop methods and tools for real time updating. A periodic recalculation (dictated by the triggers) currently allows for the updating of data cubes. This method is insufficient however for emergency decisions that require up-to-date information at all times. Thanks to the incredible evolution in the power of processors and the lower costs for RAM, the possibility for producing ***geospatial data cubes on demand*** exists. However, additional challenges await the processing of geospatial data included in these cubes. Such a solution currently does not exist, because it cannot be assumed by the traditional functions of

---

[103] *The mosaic structure is a data structure that enables the representation of all the space covered and the associated storage of data, by breaking it down into surfaces or cells considered to be homogeneous. The most current mosaic structures are the matrix structure and the quaternary data structure (Dictionnaire terminologique de l'office de la langue française, 2007).*

non-spatial technologies and it involves concepts of automatic generalization and complex multiple representations.

Moreover, the developers must pay more and more attention to the interfaces and functions used in order to keep their utilization simple. That is why **the improvement of decision-support tools** using geospatial data cubes is also an important research focus. For example, it becomes necessary to increase the usability of the SOLAP user interface for decision-makers, which will facilitate their insertion in a dashboard; to enrich spatial analysis functions for SOLAP analysts; to couple new technologies with the tools already used by specialists such as SAS, SPSS, SAP, PeopleSoft; to automatically insert contextual warnings, and to help in evaluating the best data sources and the best ETL processes for populating the cubes. Also, with the development of web portals, the decision-support tools of tomorrow should no longer be associated to particular software solutions, but rather to open-source solutions, data sharing protocols and official standards.

Although the definition of spatial algorithms for ***spatial data mining*** is still pertinent today, a more formal integration of the OLAP and data mining would also be desirable. The execution of topological and metric spatio-temporal operations can advantageously be executed in spatial data using ***spatial data mining*** tools and the results obtained stored in the cube. The results could then be analyzed via OLAP-type operations (e.g. drill-down, drill-up).

Other research domains in geomatics, such as mobile GIS, location-based services (LBS), collaborative GIS, geovisualization (Adrienko et Adrienko, 2007), geographic hypermedia systems (Williams *et al.*, 2006), spatio-temporal systems, and geosimulation systems are already of interest for the health community and their integration with decision-support technologies is currently in progress within university laboratories.

Lastly, the problems encountered, even if technically more complex than those of GIS and DBMS transactional systems, are of the same nature and consequently geo-analytical projects can be completed successfully right now. However, problems specific to health such as constraints related to the instability of statistical rates on smaller regions and the confidentiality of data will still subsist within analytical tools (e.g. SOLAP) although reduced by the use of aggregated data.

Despite all the benefits of new technologies, it is still possible that certain resistances or scientific motivations can persist and delay the adoption of these solutions in the health world although the approach is promising.

7. CONCLUSION

This literature review presents an analysis of the geospatial applications developed currently in the health field. What has been revealed is that these applications are mostly deployed using GIS technologies and web mapping technologies, but that more and more teams are working to extend the capabilities of these applications in order to support analytical needs. However, this study also presents the possibilities offered by the fields of computer science and analytical geomatics. With an innovative vision, this report presents, to environmental health professionals and politicians, the interest to go beyond the classical geographic information systems and web mapping systems to better support decision-making and the discovery of knowledge in environmental health. The presentation of a new approach (i.e. multidimensional) will now encourage the deployment of decision-support applications using innovative technologies (e.g. spatial OLAP, spatial dashboards and spatial data mining), some of which have recently reached a level of maturity allowing for their deployment in an operational environment. The decision-support applications can be developed in an autonomous manner or in addition to the GIS and DBMS in order to increase the value of the data stored in these technologies. The cost associated to the deployment of these new technologies are in fact a fraction of the cost of the transactional systems because they typically aim at better exploiting the data that are already available in the organization. When transactional data are of excellent quality and when using mature spatial decision-support technologies, efforts required to deploy the new analytical application are reduced.

Furthermore, a proper data cube design allows for working with aggregated data even if the detailed data are incomplete or of variable quality (see for example Miquel *et al.*, 2002). The design phase of the data cube is thus essential in such cases and can even lead to the inclusion of contextual warnings in the spatial decision-support application, to the implementation of adapted user training programs, or to the implementation of specific data access permissions (Levesque *et al.*, 2007).

The research work related to these tools in the field of analytical geospatial systems is increasing. A new scientific and professional community has developed itself particularly over the last 3 years around these technologies. Certain members of this community collaborate with epidemiologists and other environmental health professionals. Current research on geospatial databases is articulated around various aspects including the improvement of methods and tools for the design of geospatial data cubes, their implementation, the evaluation of their quality, the improvement of decision-support tools that use geospatial data cubes (e.g. second generation of SOLAP tools), and the deployment of location-based services. The emergence of mobile, real time and location-based aspects in the analytical geomatics field leads us to believe that the future of these technologies is promising and will result in new types of services to health specialists.

8. REFERENCES

Adrienko, G., N., Adrienko, 2007, Multimodal analytical visualization od spatio-temporal data, In: *Multimedia Cartography*, W. Cartwright, MP Paterson, G. Gartner (eds), 2nd Edition, Springer, pp. 327-346.

Arleth, M., 1999, Problems in screen map design, *19th International Proceedings Ottawa ICA* 1999.

Artnick, B., G. Vidmar, J., Javornik, 2002, OLAP for health statistics : How to turn a simple Spreadsheet into a powerful analytic tool, *12th Statistical Days – Proceeding*, Radenci, Slovenia, Statistical Society of the Republic of Slovenia, pp. 476-481. http://www.mf.uni-lj.si/ibmi/biostat-center/predtiski/Radenci_Artnik_Vidmar_Javornik_OLAPExcel.pdf.

Aufaure, M.-A., L. Yeh, K. Zeitouni, 2000, chapitre 1. Fouille de données spatiales dans le temps, l'espace, et l'évolutif en sciences du traitement de l'information, H. Prade, R. Jeansoulin, C. Garbay, (Eds.), Cépaduès, Toulouse, 2000. 458 pages.

Badard, T., P. Gosselin, Y. Bédard, J. Pouliot, 2006, Un premier outil web interactif pour mieux comprendre les vulnérabilités de santé liées au climat, *Géomatique 2006*, 25-26 octobre 2006, Montréal, Canada.

Bédard, Y., 1997, Spatial OLAP, Paper presented at the Annual Forum on R&D, *Geomatics VI*, Canadian Institute of Geomatics, Montreal, November, Canada.

Bédard, Y., 2005, Integrating GIS and OLAP: a New Way to Unlock Geospatial Data for Decision-making, *Directions on Location Technology and Business Intelligence*, May 2-4th, Philadelphia, United-States.

Bédard, Y., P. Gosselin, S. Rivest, M.-J. Proulx, M. Nadeau, G. Lebel, M.F. Gagnon, 2003, Integrating GIS components with knowledge discovery technology for environmental health decision support, *International Journal of Medical Informatics*, No. 70, Vol. 1, pp. 79-94.

Bédard, Y., S. Larrivée, M.-J. Proulx, F. Létourneau, P.-Y. Caron, 1997, *Étude de l'état actuel et des besoins de R&D relativement aux architectures et technologies des data warehouses appliquées aux données spatiales*, Rapport de recherche remis au CRDV, 98 p., Mars.

Bédard, Y., T. Merrett , J. Han, 2001, Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, In : *Geographic Data Mining and Knowledge Discovery*, Taylor & Francis, Vol. Research Monographs in GIS, Chap. 3, pp. 53-73.

Bédard, Y., M.-J. Proulx & S. Rivest, 2005, Enrichissement du OLAP pour l'analyse géographique : exemples de réalisation et différentes possibilités technologiques, *Revue des Nouvelles Technologies de l'Information - Entrepôts de données et l'Analyse en ligne*, sous la direction de F. Bentayeb, O. Boussaïd, J. Darmont et S. Loudcher, Cépaduès-Éditions, France, pp. 1-20.

Bédard, Y., S. Rivest, M.-J. Proulx, 2007, Merging GIS and Business Intelligence for Spatial Data, In: Karimi, H.A (Ed(s)), *Encyclopedia of Geoinformatics*, Idea Group Publishing. Accepted for publication in March 2007.

Bédard, Y., S. Rivest, M.-J. Proulx, 2007, Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective, In: Robert Wrembel & Christian Koncilia (Ed(s)), *Data Warehouses and OLAP: Concepts, Architectures and Solutions*, Chap. 13, IRM Press (Idea Group), London, UK, pp. 298-319.

Bell, S., R.E. Hoskins, L.W. Pickel, D. Wartenberg, 2006, Current practices in spatial analysis of cancer data: mapping health statistics to inform policymakers and the public, *International Journal of Health Geographics 2006*, No 5, Vol 49.

Berndt, D.J., A. R. Hevner, J. Studnicki, 2003, The Catch data warehouse: support for community health care decision-making, *Decision Support Systems*, ELSEVIER, Vol. 35, p. 367-384.

Berndt D. J., J. W. Fisher, A. R. Hevner, J. Studnicki, 2001, Healthcare Data Warehousing and Quality Assurance, *Computer*, vol. 34, no. 12, pp. 56-65.

Bhowmick, T, A. Gruver, A.C. Robinson, A.M. MacEachren, E. Lengerich, 2006, Using e-Delphi to Evaluate the Pennsylvania Cancer Atlas, *Proceedings AutoCarto 2006*, Vancouver, WA, June 26-28.

British Columbia Centre for Disease Control, 2007, *Interactive GIS, Mapping for West Nile Virus*, Internet site accessed November 5th, 2007, http://maps.bccdc.org.

Boulos, M.N.K., 2004, Towards evidence-based, GIS-driven national spatial health information infrastructure and surveillance services in the United Kingdom, *International Journal of Health Geographics 2004*, Vol. 3, No. 1. 50 pages.

Boulos, M.N.K., A.V. Roudsari, E.R. Carson, 2001, Mythological review, Health Geomatics: An Enabling Suite of Technologies in Health and Healthcare, *Journal of Biomedical Informatics*, Vol. 34, pp. 195–219.

Buckeridge, D.L., R. Mason, A. Robertson, J. Frank, R. Glazier, L. Purdon, C.G. Amrhein, N. Chaudhuri, E. Fuller-Thomson, P. Gozdyra, D. Hulchanski, B. Moldofsky, M. Thompson, R. Wright, 2002, Making health data maps: a case study of a community/university research collaboration. *Social Science & Medicine* Vol. 55, No. 7, pp.1189-1206.

Buzan, T., B. Buzan, 2003, *Mind Map, dessine-moi l'intelligence*, 2è édition. Paris: Éditions de l'Organisation, seconde édition, 328 pages

Caron, P.Y., 1997, *Le potentiel de l'approche OLAP pour améliorer les capacités d'analyses spatio-temporelles dans les SIG*, mémoire de maîtrise, Département des sciences géomatiques, Université Laval.

Chung, K., D.H. Yang, R. Bell, 2004, Health and GIS: Toward Spatial Statistical Analyses, *Journal of Medical Systems 2004*, Vol. 28, pp. 349-360.

Codd, E.F., S.B. Codd, C.T. Salley, 1993, *Providing OLAP user-analysts: An IT mandate*, Arbor software paper (www.arboresoft.com/papers), 23 pages.

Cromley, H.K. 2003, GIS and diseases, *Annual Review of Public Health 2003*, Vol. 24, pp.7–24.

Da Sylva, J., V.C. Times, A.C. Salgado, 2006, An Open Source and Web Based Framework for Geographic and Multidimensional Processing, *SAC'06*, April 23-27, 2006, Dijon, France.

De Lepper, M.J., H.J. Scholten, R.M. Stern, 1995, *The added value of Geographical Information Systems in Public and Environmental Health*, Fisrt Edition, Kluwer Academic Publishers, Netherlands, 370 pages.

Devlin, B, 1997, *Data Warehouse from Architecture to Implementation*, Addison Wesley Longman inc., 432 pages.

Eckerson, W, 2005, What are Performance Dashboards? *DM Review Magazine*, November.

Elliott, P., D. Wartenberg, 2004, Spatial epidemiology: current approaches and future challenges, *Environmental Health Perspectives 2004*, Vol. 112, No.9, pp.998–1006, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1247193.

Elliott, S.J., J. Eyles, P. De Luca, 2001, Mapping health in the Great Lakes areas of concern: a user-friendly tool for policy and decision makers. *Environmental Health Perspectives, Vol.* 109, December, Supplement 6, pp. 817-826.

Fayyad, U., G. Piatetsky-Shapiro, P. Smyth, 1998, From data mining to knowledge discovery: an overview. *Advances in Knowledge Discovery and Data Mining, MIT Press*, Vol. 1, pp. 1–36.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth, 1996, Knowledge discovery and data mining: Towards a unifying framework. In: Proc. 2nd International Conference on Knowledge Discovery and Data Mining KDD, pp. 82-88, Portland, Oregon.

Ferland, Y., 2006, *Le 100e anniversaire de l'Atlas du Canada, 1906-2006*, Cahiers de géographie du Québec, Volume 50, numéro 140, septembre, pp. 263-264. Disponible pour téléchargement à partir de : http://www.cgq.ulaval.ca/textes/vol_50/no_140/12-Ferland.pdf

Fernandez, A., 2000, *Les nouveaux tableaux de bord des décideurs*, Éditions d'Organisation, Paris, 451 pages.

Fernandez, A., 2005, *L'essentiel du tableau de bord*, Éditions d'Organisation, Paris, 178 pages.

Fernandez, A., 2007, *Piloter la performance*, Internet site accessed February 2007, http://www.piloter.org/

Fortin, C., R. Rousseau, 1989, *Psychologie cognitive: une approche de traitement de l'information*. Presses de l'Université du Québec, 434 pages.

Franklin, C., P. Hane, 1992. An Introduction to Geographic Information Systems: Linking Maps to Databases. *Database*, Vol.5, No.2, April, pp. 12-22.

Freier, J., 2000, Mapping Outbreaks Using GIS. In *Proceedings of Outbreak*, Symposium at American Veterinary Medical Association, 23 July, 2000, Schaumburg, http://www.aphis.usda.gov/vs/ep/avma/freier.html

Gill, H.S., P.C. Rao, 1996, *The Official Client/Server Computing Guide to Data Warehouse*, QUE Corporation, 382 pages.

Goodchild, M. F., 1992, Geographical data modeling, *Computers and Geosciences*, Vol. 18, pp. 401-408.

Gosselin, P., Y. Bédard, M. Jerrett, S. Elliott, R. Catelan, P. Poitras, A. Gingras, 2000, *GIS and OLAP in Health Surveillance: Needs Analysis for Successful Integration*, Final Report for National Health Surveillance Infrastructure, Health Protection Branch, Health Canada, CHUQ, February, 70 pages.

Gosselin, P. G. Lebel, S. Rivest, M. Douville-Fradet, 2005, The Integrated System for Public Health Monitoring of West Nile Virus (ISPHM-WNV): a real time GIS for surveillance and decision-making, *International Journal of Health Geographics, Vol.* 4, p. 21.

Gouvernement du Québec – Ministère de la santé et des services sociaux, 2004, *Plan intégré des ressources informationnelles en santé publique*, Publié par la direction générale de la santé publique.

Han, J., M. Kamber, 2001, Data Mining: Concepts and Techniques. Morgan Kaufmann.

Han, J., K. Koperski, N. Stefanovic, 1997, GeoMiner: A System Prototype for Spatial Data Mining. In Proceedings of the 1997 *ACM-SIGMOD International Conference on Management of Data*, Tucson, Arizona, May, pp. 553-556.

Han, J., N. Stefanovic, K. Koperski, 1998, Selective Materialization: An Efficient Method for Spatial Data Cube Construction, Research and Development in Knowledge Discovery and Data Mining. In: Proceedings of the *Second Pacific-Asia Conference, PAKDD'98*, pp. 144-158.

Han, J., Y. Fu, W. Wang, J. Chiang, W. Gong, K. Koperski, D. Li, Y. Lu, A. Rajan, N. Stefanovic, B. Xia, O.R. Zaiane, 1996, DBMiner: A System for Mining Knowledge in Large Relational Databases, Proc. 1996 *Int. Conf. on Data Mining and Knowledge Discovery (KDD'96)* , Portland, Oregon, pp. 250-255.

Hristovski D, M. Rogac, M. Markota, 2000, Using data warehousing and OLAP in public health care. *Proc AMIA Symp*, pp. 369–373.

Inmon, W.H., 1996a, *Building the Data Warehouse*, John Wiley & son, 410 pages.

Inmon, W.H., 1996b, The data warehouse and data mining, Communications of the *ACM [0001-0782]*, vol.39 no.11, pp. 49-50.

Institut National de Santé Publique du Québec, 2004, *La géomatique l'infocentre de santé publique du Québec: Stratégie de géomatisation*, Juin 2006, 24 pages.

Jaishankar, R., C.P. Jhonson, 2006, Geomatics and Public Health, *Indian Journal of Public Health*, Vol. 50, No.1, January-March 2006, pp. 24-27.

Kimball, R., 1997, A dimensional Modeling Manifesto, *DMBS Magazine*, August, 1997, Vol. 10, No. 9, p. 59.

Kimball, R., 2002a, *The Data Warehouse Toolkit*, John Wiley and Sons, Inc, Second edition, 436 pages.

Kimball, R, 2002b, Two Powerful Ideas, The foundations for modern data warehousing, *Intelligent Enterprise Magazine*, September, http://www.intelligententerprise.com//020917/515warehouse1_1.jhtml.

Kraak, M.J., A. Brown, 2001, *Web Cartography - Developments and prospects*, Taylor & Francis, New York, 213 pages.

Levesque, M.-A., Y. Bédard, M. Gervais, R. Devillers, 2007, Towards a safer use of spatial datacubes: communicating warnings to users, Proceedings of the *5th International Symposium on Spatial Data Quality*, June 13-15, Enschede, The Netherlands.

LGS Group, 2000, *Analysis of Health Surveillance Business Intelligence Tools & Applications : Analysis Methodology and Criteria*, project report, 19 pages.

MicroStrategy, 2006, *Summary Results from the OLAP Survey 6*, 20 pages.

Miller, H., J. Han, 2001, *Geographic Data Mining and Knowledge Discovery*, London, Taylor & Francis, 372 pages.

Miquel, M., Y. Bédard, A. Brisebois, J. Pouliot, P. Marchand, J. Brodeur, 2002, Modeling Multidimensional Spatio-temporal Data Warehouse in a Context of Evolving Specifications, Joint *Int. Symp. International Society for Photogrammetry and Remote Sensing (ISPRS) Commission IV*, SDH 2002, 95th Annual CIG Conference, July, 8-12.

Mowat, D., P. Gosselin, Y. Bedard, J. Moore, M. Goddard, 2000, Improving health surveillance in Canada - What are the needs? Victoria, BC, Canada.

New Brunswick Lung Association, 2006, *Canadian Geospatial Data Infrastructure, Gaps and opportunities in public health*, rapport, New Brunswick, 145 pages.

Office de la langue française, Québec, 2007, Internet site accessed February 2007, http://www.granddictionnaire.com/btml/fra/r_motclef/index1024_1.asp

Open GIS Consortium, 2007, Glossary of terms, http://www.opengeospatial.org/ogc/glossary/, Accessed November 5th, 2007.

Peterson, Michael (ed.), 2003, Maps and the Internet, Elsevier, 468 pages.

Proulx, M.-J., Y. Bédard, 2004a, Le potentiel de l'approche multidimensionnelle pour l'analyse de données géospatiales en comparaison avec l'approche transactionnelle des SIG, *Colloque Géomatique 2004 - Un choix stratégique!* Montréal, 27-28 octobre.

Proulx, M.-J., Y. Bédard, 2004b, *L'émergence du SOLAP en support au cycle décisionnel : Défis et enjeux à la recherche pour la défense- Valcartier*, présentation, 31 mars 2004, 94 pages.

Proulx, M.-J., M. Nadeau, S. Rivest, Y. Bédard, P. Gosselin, G. Lebel, 2002, Outil d'analyse multidimensionnel spatial (SOLAP): ICEM/SE appliqué à la surveillance en santé environnementale, *Colloque AGMQ*, 8-oct., Québec, Canada.

Radding, A., 1995, Warehouse wake-up call - Building a better Warehouse the dream may be appealing, but in reality building a data warehouse can turn into a difficult task. *InfoWolrd*, November 20, Vol. 17, Issue 47.

Rawling J., H. Kucera, 1997, Trial and tribulations of implementing a spatial data warehouse, 11th Annual Symposium on Geographical Information Systems, Vancouver, pp.510-513.

Reichenbacher, T., 2007, Adaptation in mobile and ubiquitous cartography, In: *Multimedia Cartography*, W. Cartwright, M.P. Peterson, G. Gartner (Eds), 2nd edition, Springer, pp 383-397.

Rezaeian, M., G. Dunn, S. St Leger, L. Appleby, 2007, Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary, *Journal of Epidemiology Community Health*, Vol. 61, pp.98–102.

Rivest, S., 2000, *Investigation des modes d'intégration physique entre un serveur de base de données multidimensionnelle et un SIG*, essai de M.Sc., Faculté de Foresterie et de Géomatique, Université Laval, 84 pages.

Rivest, S., Y. Bédard, M.-J. Proulx, M. Nadeau, 2003, SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis. In: Proceedings of the *ISPRS Joint Workshop on Spatial, Temporal and Multi-Dimensional Data Modelling and Analysis*, Quebec, Canada.

Robinson, A.C., J. Chen, G. Lengerich, H. Meyer, A.M. Mac Eachren, 2005, Combining usability techniques to design geovisualization tools for epidemiology. *Cartography and Geographic Information Science*, Vol. 32, No. 4.

Rivest, S., Y. Bédard, P. Marchand, 2001, Towards better support for spatial decision-making: defining the characteristics of Spatial On-Line Analytical Processing. *Geomatica*, Vol. 55, No. 4, pp. 539-555.

Rushton, G., 2003, Public Health, GIS, and Spatial Analytic Tools, Annual Review Public Health 2003, Vol. 24, pp. 43-56.

Scotch, M., B. Parmanto, 2005, SOVAT: Spatial OLAP Visualization and Analysis Tool, Proceedings of the *38th Hawaii International Conference on System Sciences – 2005*, 7 pages.

Scotch, M., B. Parmanto, C.S. Gadd, R.K. Sharma, 2006, Exploring the role of GIS during community health assessment problem solving: experiences of public health professionals, *International Journal of Health Geographics 2006*, Vol. 5, No. 39, 10 pages.

Shuai, J., P. Buck, P. Sockett, J. Armini, F. Pollari, 2006, A GIS-driven integrated real time surveillance pilot system for national West Nile virus dead bird surveillance in Canada, *International Journal of Health Geographic 2006*, Vol. 5, No. 17.

Sonesson, C., D. Bock, 2003, A review and discussion of prospective statistical surveillance in public health. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 166, No. 1, pp. 5-21.

Standing, L., 1973, Learning 10000 Pictures, *Quaterly Journal of Experimental Psychology*, Vol. 25, pp. 207-222.

Stefanovic, N., 1997, Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data. *Master's thesis*, Simon Fraser University, Canada.

Stefanovic, N., J. Han, K. Koperski, 2000, Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. *IEEE Transactions on Knowledge Discovery and Data Engineering*, Vol. 12, No. 6, pp. 938-958.

TDWI, 2006, Cancer care Ontario: Creating one-step access to cancer services information, *What works*, Volume 22, November 2006, http://www.dw-institute.com/Publications/WhatWorks/display.aspx?id=8178.

Thearling, K., 2007, *An Introduction to Data Mining - Discovering hidden value in your data warehouse*, http://www.thearling.com/text/dmwhite/dmwhite.htm.

Thomsen, E., 2002. *OLAP Solutions: Building Multidimensional Information Systems,* Second Edition, New York, John Wiley, 688 pages.

Turban, E.,J.E. Aronson, T.P. Liang, 2005, *Decision Support Systems and Intelligent systems,* 7th edition, Prentice Hall, 936 pages.

Vaitis, M., M. Tzagarakis, 2005. Towards the Development of Open Cartographic Hypermedia Systems, *Lecture Notes in Geoinformation and Cartography, Geographic Hypermedia: Concepts and Systems,* pp.211-231.

Vanasse A., J. Courteau, M. Plante, G. Orzanco, M. Courteau, C. Dean, P. Dagenais, J.-P. Grégoire T. Niyonsega, 2007, The SIST-MC project: Technological Model, 9th GEOIDE Annual Scientific Conference, Halifax, June 6th-8th.

Voyer, P., 2000, Tableaux de bord de gestion et indicateurs de performance, Presse de l'Université du Québec. 439 pages. http://books.google.com/books?id=pP1P67D-PJMC&pg=PR3&ots=pExWNTMBSA&dq=Tableaux+de+bord+de+gestion+et+indicateurs+de+performance&sig=qpLoOg3r256yZ9ViAhG40DXNYjY.

Williams, P., E. Siekierska, C. Armenakis, F. Savopol, C. Siegel, J. Webster, 2006, Visualization and Hypermedia for Decision Making, in *Geographic Hypermedia: Concepts and Systems*, E. Stefanakis, M. Peterson, C. Armenakis, V. Delis (Eds), Springer-Verlag Berlin Heidelberg, pp. 309-328.

Yang, C., D.W. Wong, R. Yang, M. Kafatos, Q. Li, 2005, Performance-improving techniques in web-based GIS, International Journal of Geographical Information Science, Vol. 19, No. 3, pp. 319-342.

Youngworth, P., 1995, OLAP, *Database advisor*, December, p. 38.

Zeitouni, K., 2006, Fouille de données complexes, notes de cours en ligne sur : http://www.prism.uvsq.fr/users/karima/DM.