

# SOLAP technology: Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data

Sonia Rivest<sup>a,1</sup>, Yvan Bédard<sup>a,\*</sup>, Marie-Josée Proulx<sup>a,1</sup>, Martin Nadeau<sup>a,1</sup>,  
Frederic Hubert<sup>a,1</sup>, Julien Pastor<sup>a,b</sup>

<sup>a</sup> *Department Geomatics Sciences, NSERC Industrial Chair in Geospatial Databases for Decision Support, Centre for Research in Geomatics, Laval University, Quebec, Canada, G1K 7P4*

<sup>b</sup> *KHEOPS Technologies, Montreal, Canada (now at Leica Geosystems, France)*

Received 4 October 2004; received in revised form 6 October 2005; accepted 12 October 2005

Available online 17 November 2005

## Abstract

To support their analytical processes, today's organizations deploy data warehouses and client tools such as OLAP (On-Line Analytical Processing) to access, visualize, and analyze their integrated, aggregated and summarized data. Since a large part of these data have a spatial component, better client tools are required to take full advantage of the geometry of the spatial phenomena or objects being analyzed. With this regard, Spatial OLAP (SOLAP) technology offers promising possibilities. A SOLAP tool can be defined as "a type of software that allows rapid and easy navigation within spatial databases and that offers many levels of information granularity, many themes, many epochs and many display modes synchronized or not: maps, tables and diagrams" [Bédard, Y., Proulx, M.J., Rivest, S., 2005. Enrichissement du OLAP pour l'analyse géographique: exemples de réalisation et différentes possibilités technologiques. In: Bentayeb, F., Boussaid, O., Darmont, J., Rabaseda, S. (Eds.), *Entrepôts de Données et Analyse en ligne, RNTI B.1*. Paris: Cépaduès, pp. 1–20]. SOLAP tools offer a new user interface and are meant to be client applications sitting on top of multi-scale spatial data warehouses or datacubes. As they are based on the multidimensional paradigm, they facilitate the interactive spatio-temporal exploration of data. The purpose of this paper is to discuss how SOLAP concepts support spatio-temporal exploration of data and then to present the geovisualization, interactivity, and animation features of the SOLAP software developed by our research group. This paper first reviews the general concepts behind OLAP and SOLAP systems. This is followed by a discussion of how these SOLAP concepts support spatio-temporal exploration of data. In the subsequent section, SOLAP software is introduced along with features that enable geovisualization, interactivity and animation. © 2005 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

**Keywords:** spatial decision-support; geovisualization; multidimensional databases; Spatial On-Line Analytical Processing (SOLAP); Geographic Information Systems (GIS)

\* Corresponding author. Tel.: +1 418 656 2131; fax: +1 418 656 3607.

E-mail addresses: [sonia.rivest@scg.ulaval.ca](mailto:sonia.rivest@scg.ulaval.ca) (S. Rivest), [yvan.bedard@scg.ulaval.ca](mailto:yvan.bedard@scg.ulaval.ca) (Y. Bédard), [marie-josée.proulx@scg.ulaval.ca](mailto:marie-josée.proulx@scg.ulaval.ca) (M.-J. Proulx), [martin.nadeau@scg.ulaval.ca](mailto:martin.nadeau@scg.ulaval.ca) (M. Nadeau), [frederic.hubert@scg.ulaval.ca](mailto:frederic.hubert@scg.ulaval.ca) (F. Hubert), [julien.pastor@scg.ulaval.ca](mailto:julien.pastor@scg.ulaval.ca), [julien.pastor@leica-geosystems.com](mailto:julien.pastor@leica-geosystems.com) (J. Pastor).

<sup>1</sup> Tel.: +1 418 656 2131; fax: +1 418 656 3607.

## 1. Introduction

Organizations collect huge amounts of data within their day-to-day operations. These data are usually stored in transactional systems, which are primarily optimized to ensure consistency, efficient updates, secured concurrent accesses, efficient execution of a large number of small transactions, and a near fault tolerant availability of data. They provide fast response times for SQL-type queries involving a small number of occurrences (Date, 2003; Shekhar and Chawla, 2002). Relational database management systems (RDBMS) and universal servers are the backbone of transactional databases and most GIS applications are built using such a transactional approach. However, the way transactional databases are optimized makes the data difficult to exploit by managers and analysts who need aggregated and summarized information, rapid comparisons in space and time, syntheses over millions of occurrences, trends discovery and other complex operations to support their tactical and strategic decision-making processes. The same holds for GIS and geospatial data. Since transactional systems are not designed to support the decisional processes, new types of systems have been developed to specifically fulfil decisional needs; they are called “Analytical Systems” and are known on the market as “Business Intelligence” (BI) solutions. These systems, in which the data warehouse is usually a central component, are optimized to facilitate complex analysis and to improve the performance of database queries involving thousands or more occurrences (ex. aggregated information) (Inmon, 2002). The most widely used BI solutions are OLAP (On-Line Analytical Processing) systems, which provide a unique capability to interactively explore the data warehouse.

In the BI world, data warehouses are based on data structures called “multidimensional”. The term “multidimensional” was coined in the mid-1980s by the community of computer scientists who were involved in the extraction of meaningful information from very large statistical databases (ex. national census) (Rafanelli, 2003). This concept of multidimensionality refers to neither the  $x$ ,  $y$ ,  $z$ , and  $t$  dimensions typically addressed by the GIS community nor to the multiple formats (ex. vector, raster, DTM) as considered by some GIS specialists. Section 2 will present this concept of dimensionality along with other concepts used for Spatial OLAP (SOLAP).

SOLAP has been developed to fully exploit the powerful concepts brought by the multidimensional database structure, and to add spatial extensions that provide highly interactive map visualization and data

exploration. The underlying OLAP approach supports the iterative nature of the analytical process because it allows the user to explore and navigate across the different themes (dimensions) at different levels of detail and to rapidly visualize the facts or data at the intersections of these dimensions, whatever their level of aggregation. This gives access to all the possible views, or all the possible combinations, of the data. This is typically done using a few mouse clicks, and response times are within 10 s, a characteristic that is necessary to remain within the cognitive band identified by Newell (1990) for adequate decision support. In addition, such level of interactivity facilitates the emergence of new hypotheses to solve problems and encourages knowledge discovery for scientists.

It has been estimated that up to 80% of all data stored in corporate databases may have a spatial component (Franklin, 1992). To fully exploit this component in the context of interactive spatio-temporal exploration and analysis of data, today’s commercial tools must be adapted or new tools must be built. For example, it has been shown that OLAP already possesses a certain potential to support spatio-temporal analysis (Caron, 1998). However, without a cartographic user interface to view the spatial distribution and correlations of phenomena, and without spatial operators to navigate through aggregated spatial data, the analysis may be incomplete or even lead to false conclusions in some situations. Geographic information systems (GIS) are potential candidates to support decisional needs, but despite their capabilities, it is recognized that today’s GIS are designed neither to support decisional applications nor to support highly interactive navigation through spatial data at different levels of aggregation and through different epochs. GIS are typically implemented as transactional systems and alternative solutions must be used for analytical purposes (Bédard et al., 2001). Among them, the coupling of GIS and OLAP into Spatial OLAP (SOLAP) is an interesting path to explore.

Sometimes qualified as “Keyboardless-GIS”, but offering more than a GIS in terms of navigation within multidimensional datasets, SOLAP tools possess an intuitive user interface allowing non-technical users to easily access, slice, dice, drill, swap (see Section 2.2), visualize and analyze their data.

The purpose of this paper is to discuss how SOLAP concepts support spatio-temporal exploration of data and to present the geovisualization, interactivity, and animation features of the SOLAP software developed by our research group. This paper first reviews the general concepts behind OLAP and SOLAP systems.

This is followed by a discussion of how these SOLAP concepts support spatio-temporal exploration of data. In the subsequent section, a SOLAP technology is introduced, along with features that enable geovisualization, interactivity, and animation.

## 2. Review of OLAP and SOLAP concepts

### 2.1. OLAP concepts

The term On-Line Analytical Processing, or OLAP, was coined in the early 1990s by E.F. Codd, the pioneer of relational systems, in order to clearly indicate that something different was needed for analytical processes, something different from well-known OLTP (On-Line Transactional Processing, the typical type of processing offered those days by DBMS). OLAP has been defined for the first time as “(...) the name given to the dynamic enterprise analysis required to create, manipulate, animate and synthesize information from exegetical, contemplative and formulaic data analysis models. This includes the ability to discern new or unanticipated relationships between variables, the ability to identify the parameters necessary to handle large amounts of data, to create an unlimited number of dimensions, and to specify cross-dimensional conditions and expressions” (Codd et al., 1993). The reader is referred to the aforementioned paper for a detailed description of each data analysis model. Other definitions of OLAP have been proposed, including: “a software category intended for the rapid exploration and analysis of data based on a multidimensional approach with several aggregation levels” (Caron, 1998) and “a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user” (AltaPlana, 2004). For additional information and a detailed OLAP glossary, see (AltaPlana, 2004).

OLAP technology is based on the multidimensional database approach, which introduces concepts that differ from the concepts found in the transactional database approach. The key multidimensional concepts include: dimensions, members, measures, facts and data cubes (Berson and Smith, 1997; AltaPlana, 2004; Pendse, 2000; Thomsen, 2002). The *dimensions* represent the themes of interest for a user, or the analysis axes of an  $N$ -dimensional thematic space (ex. “time”, “cancer”, “population” and “administrative zones” in a public health context). As mentioned previously, the

concept of dimension in this context is different from the concept of dimension in a spatial reference system (i.e.  $X$ ,  $Y$ ,  $Z$  axes); nevertheless, dimensions can be spatial such as those using location names solely (ex. names of country-province-area-city). Within the multidimensional database context, the dimensions are also seen as the independent variables included in the analysis. Dimensions are organized hierarchically into levels of granularity, or levels of details. An example would be the “administrative zones” dimension where we could use “province” (the most general level), “regional health authorities” (a more detailed level), “local health authorities” (the most detailed level).

A dimension contains members. For example, “1998”, “stomach cancer”, “women” and “Gaspésie region” are, respectively, members of the “time”, “cancer”, “population” and “administrative zones” dimensions. The members of one level (ex. the different months, for a time dimension) can be aggregated to constitute the members of the next higher level (ex. the different years). Dimensions can be of different types: temporal, spatial (non-cartographic in the case of a conventional (non-spatial) OLAP tool) and descriptive (or thematic) (Bédard et al., 2001).

The *measures* (ex. count, sum, standardized rate) are the numerical attributes analyzed against the different dimensions. A measure can then be considered as the dependent variable while dimension members are the independent variables. Each of the measures depends on a set of dimensions, which provide the context for the measure. The dimensions together are assumed to uniquely determine the measure (Chaudhuri and Dayal, 1997). For example, the value of a “standardized rate” measure depends on the members of the “time”, “cancer”, “population” and “administrative zones” dimensions. Measures can be based on complex formulas and can contribute to the creation of sophisticated mathematical models for use in the development of scenarios, for example.

Each potential combination of dimension members, with the resulting measure(s) value for a particular aggregation level represents a fact. For example, the *standardized rate of incidence of stomach cancer for the year 2001, for the women and for the Gaspésie region is 1.24* is a fact; here the first values are members while the last one is a measure and their combination constitutes this unique fact.

A data cube (also called hypercube when more than 3 dimensions are used) is composed of a set of measures aggregated according to a set of dimensions (Thomsen et al., 1999). Inside a data cube, possible aggregations of measures on the possible combinations of dimension

members (the facts) can be pre-computed (to a certain level) and stored to increase query performance. Several data cubes can be built from the same sources of data for different interactive exploration and analysis needs, as they usually are read-only datasets. Fig. 1 presents the multidimensional database concepts.

The common OLAP architecture usually comprises three components: the multidimensionally structured database, the OLAP server and the OLAP client that accesses the database via the OLAP server. Depending on the technology (relational, multidimensional, ...) used to implement the OLAP database, it is possible to distinguish three OLAP approaches: relational OLAP (ROLAP), multidimensional OLAP (MOLAP) or hybrid OLAP (HOLAP), which is an optimized combination of the two previous approaches (Pendse, 2000). When a relational database is used, it is possible to implement a multidimensional structure using a star, snowflake, mixed or constellation schema (Gill and Rao, 1996). Various implementation strategies can be found in Thomsen et al. (1999).

The OLAP client allows the end user to visualize the data using different types of diagrams (ex. bar charts, pie charts) and tables. It also allows the user to explore and analyze the data using different operators such as drill-down (visualize a more detailed level within a dimension), roll-up (or drill-up: visualize a more general level within a dimension), drill-across (visualize another member or another dimension at the same level of detail), swap or pivot (interchange visible dimensions or

visible and background dimensions in order to modify the content of axes used in the diagrams or tables) and slice and dice (reduce the dimensionality of the data, i.e., take a projection of the data on a subset of dimensions for selected members of the other dimensions) (Chaudhuri and Dayal, 1997).

An OLAP system is built especially to navigate within multidimensional cubes, i.e. to go from one fact to another in an interactive manner and to obtain fast responses.

## 2.2. SOLAP concepts

Following the recent implementations of spatial data warehouses, it became evident that the common client tools used to exploit non-spatial data warehouses were not sufficient to fully analyze the geometric component of the spatial data, despite the fact that the geometric component is the heart of spatial data warehouses. A new solution was then developed, which consists of combining the strengths of GIS, with the strengths of OLAP. This combination gave birth to Spatial OLAP or SOLAP, a term coined for the first time by Bédard et al. (1997).

As the architecture of an OLAP system is composed of a multidimensionally structured database, an OLAP server and an OLAP client, the architecture of a SOLAP system is composed of a multidimensionally structured spatio-temporal database, a SOLAP server and a SOLAP client (Rivest et al., 2003). The spatio-temporal database

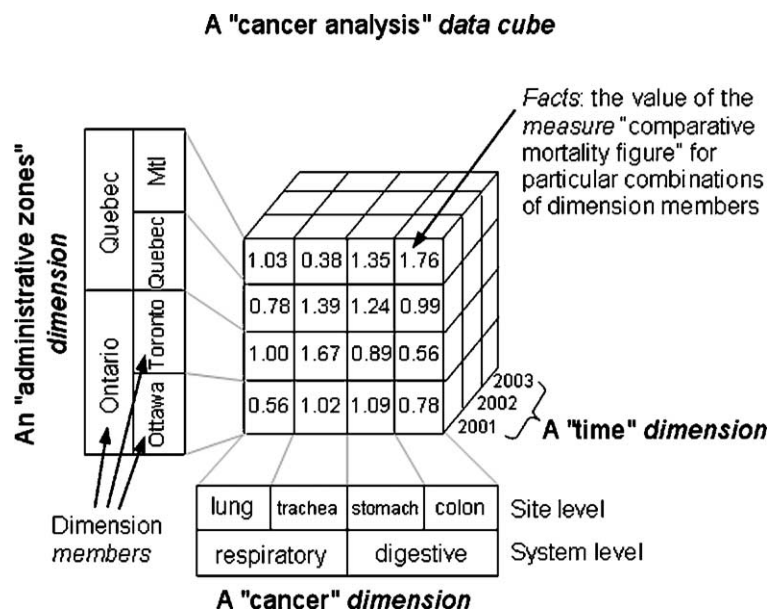


Fig. 1. A sample data cube showing the multidimensional database concepts.

stores the geometry associated with dimension members and measures (see next paragraphs for definitions). The SOLAP server handles the spatio-temporal multi-dimensional database and the numerical and spatial calculations necessary to compute the measure values associated with possible combinations of dimension members. Currently, no such server is available on the market; it must then be implemented using a custom combination of technologies. The SOLAP client can be defined as a category of software that allows navigation (i.e. to go from one fact to another) within spatial databases and that offers many levels of information granularity, many themes, many epochs, and many display modes, synchronized or not: maps, tables and diagrams (Bédard et al., 2005).

A SOLAP system supports three types of spatial dimensions (Bédard et al., 2001): the non-geometric spatial dimensions, the geometric spatial dimensions and the mixed spatial dimensions. The non-geometric spatial dimensions use nominal spatial references, i.e. only the names of places such as Canada, Province of Quebec and Quebec City. This type of spatial dimension is the only one supported by conventional (non-spatial) OLAP tools. When used within SOLAP tools, this type of spatial dimension is treated like the other descriptive dimensions and the geometric data allowing for the representation of the dimension members on maps is not used. In this case, the spatio-temporal analysis can be incomplete and certain spatial relations or correlations between the phenomena under study can be missed by the analyst. The two other types of spatial dimensions aim at minimizing this potential problem. To do so, the geometric spatial dimensions comprise, for all dimension members, at all levels of detail, geometric shapes (ex. polygons to represent country boundaries) that are spatially referenced to allow their dimension members (ex. Canada) to be visualized and queried cartographically. The mixed spatial dimensions comprise geometric shapes for a subset of the levels of details. Fig. 2 presents an example of the three types of spatial dimensions.

Each level of spatial dimensions that have geometric shapes associated to their dimension members can sup-

port spatial drilling (see next paragraphs) of cartographic features, thus increasing the number of degrees of freedom for interactive spatio-temporal exploration of data. Within a SOLAP tool, maps are used to display the members of the geometric or mixed spatial dimensions, using visual variables that relate to the values of the different measures contained in the data cube being analyzed.

A SOLAP system also supports two types of spatial measures as well as spatial dimensions. A first type of spatial measure is the set of all the geometries representing the spatial objects corresponding to a particular combination of dimension members (it is possible to have many geometric spatial dimensions). It consists of a set of coordinates, which requires a geometric operation, such as a spatial union, a spatial merge or a spatial intersection, to be computed (Han et al., 1998; Rivest et al., 2001; Stefanovic, 1997). A second type of spatial measure results from the computation of spatial metric or topological operators. Examples of this type of spatial measure could be “surface” and “distance” (Rivest et al., 2001) as well as “number of neighbours”. To implement the first type of spatial measures, it may be necessary to use pointers (stored within the multidimensional data structure) to the geometric shapes stored in another structure or software (Han and Kamber, 2001; Stefanovic, 1997). Similarly, spatial dimensions can also present the results of spatial analysis but in a hierarchical manner (ex. adjacent–adjacent by points–adjacent by only one point) and can be used to find the facts that correspond to the selected spatial operator member (Marchand et al., 2004).

The measure values are calculated by the OLAP or SOLAP server that aggregate and physically store them according to the possible combinations of dimension members. Some commercial OLAP servers materialize every possible aggregation (called *cuboids*), some other materialize no aggregation (everything is aggregated on the fly into *virtual cubes*) and some materialize only a part of the possible aggregations and use various algorithms to select the optimal aggregations to compute. In the case of SOLAP servers,

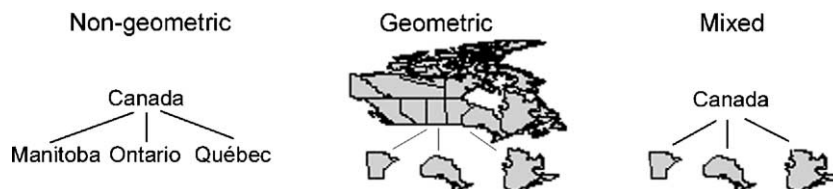


Fig. 2. The three types of spatial dimensions supported in SOLAP tools. From Rivest et al. (2003).

however, it is almost impossible to materialize all the possible geometric aggregations of spatial measures as it will result in an explosion of the necessary storage space. Algorithms have been defined in order to optimally select the spatial aggregations to be materialized. See (Han et al., 1998; Papadias et al., 2002; Shekhar et al., 2001; Stefanovic et al., 2000) for examples of such algorithms.

The measure values (spatial and non-spatial) that result from the combinations of dimension members (spatial and non-spatial) are visualized using a SOLAP client. A SOLAP client can be used with any type of SOLAP architecture: ROLAP (with or without an OLAP server), HOLAP or MOLAP. It is also possible to see SOLAP as a new type of user interface for multi-scale GIS and web mapping in order to facilitate data access. Hence, two levels of use are possible: access and analysis.

In a SOLAP client interface, variants of the OLAP operators (described in Section 2.1) are used in order to take advantage of the spatial multidimensional data structure and of the different levels of detail of the data. The operators are drill-down, roll-up (or drill-up), drill-across, swap (or pivot) and slice and dice. These SOLAP operations are available in the different types of displays (maps, diagrams or tables) and can be specialized according to the type of dimension they manipulate. *Thematic drill-down*, *thematic roll-up* and *thematic drill-across* allow to navigate from one thematic level of detail to another inside a thematic (or descriptive) dimension, while keeping the same level of spatial and temporal granularities. They can be executed directly by clicking on the elements (dimension members) of the non-cartographic displays (diagrams or tables). When defined to manipulate the data contained in the geometric or mixed spatial dimensions, the drill operators can be named *spatial drill-down*, *spatial roll-up* and *spatial drill-across*. They allow the navigation from one geometric level of detail to another within a spatial dimension, while keeping the same thematic and temporal granularities, and they can be executed directly by clicking on the elements (dimension members) shown on the maps. Similarly, *temporal drill-down*, *temporal roll-up* and *temporal drill-across* allow to navigate from one temporal level of detail to another inside a temporal dimension, while keeping the same level of spatial and thematic granularities. They can be executed directly by clicking on the elements (dimension members) of the non-cartographic displays (diagrams or tables) or by manipulating interactive time-related interface components, like a drillable interactive timeline (Pastor, 2004).

### 2.3. Categories of SOLAP

According to LGS Group (2000), various basic approaches to GIS and OLAP tool integration can be considered: GIS-dominant, OLAP-dominant and total integration (SOLAP technology). The first approach offers full GIS capabilities and a GIS graphical user interface, but only simplified access to OLAP data sources is offered and practically no OLAP functionality. The second approach is the opposite; it offers full OLAP capabilities and an OLAP graphical user interface, but limited GIS functionalities (usually only cartographic display capabilities). Implementing a SOLAP system without using SOLAP technology requires major development to implement a custom front-end for every application, and to use OLAP and GIS as back-end services. The time required to implement a functional system is reduced by an order of magnitude by using a fully integrated SOLAP technology such as the one we propose in Section 3.

### 2.4. SOLAP as an intuitive and efficient analysis tool

In the context of interactive spatio-temporal exploration and analysis of data, maps and graphics do more than make data visible; they are active instruments in the end-users thinking process (MacEachren and Kraak, 2001). Without a cartographic display, OLAP tools lack an essential feature, which could help the completion of spatio-temporal exploration and analysis processes. This geovisualization feature is present in SOLAP tools and it allows for better presentation and visualization of the data, improved dissemination and communication, enhanced analysis and better support for decision-making as implicit spatial relationships between phenomena rapidly become explicit and visually evident and new relations are more likely to emerge in the user's mind (Bédard et al., 2003). Geographic visualization, or geovisualization, can be defined as a private activity in which unknowns are revealed in a highly interactive environment. Thus, geovisualization is not a passive process of either seeing or reading maps. It is an active process in which an individual engages in sorting, highlighting, filtering, and otherwise transforming data in a search for patterns and relationships (MacEachren, 1994a).

The cartographic display of a SOLAP tool allows one to easily determine clusters, correlations and other spatial relationships that are not constrained by predefined territorial limits and that cannot be seen when using only a nominal spatial reference as supported in non-spatial OLAP tools.

Using a SOLAP tool, a user has the ability to conduct the analysis without having to master a query language or to know and understand the underlying structure of the database (Marchand et al., 2004), which may be very complex in the case of spatio-temporal databases. With a SOLAP tool, the analyst focuses on the results of the analysis rather than on the procedures required by the tool to compute the analysis results. Also, fast answers to complex queries are possible because data are (all or partly) pre-aggregated, thus reducing computation times when querying. The response times of SOLAP tools are usually included in Newell's cognitive band (Marchand et al., 2004; Newell, 1990), i.e. within 10 s. This offers a certain guarantee that the usability and performance of SOLAP do not interfere with the user's train of thoughts during data exploration and analysis and allows knowledge construction. Knowledge construction, within the geospatial sphere, is considered a developmental process, with meaning being progressively constructed and refined through a series of pre-processing and interpretative steps (Gahegan and Brodaric, 2002).

Because they are based on the multidimensional database approach, SOLAP tools offer intrinsic support for essential principles and processes of human cognition such as categorization, hierarchies and information processing (Edwards, 2001; Marchand et al., 2004). Cognitively, humans reduce the vast amount of knowledge by grouping it into categories according to their own stored knowledge (Mennis et al., 2000). Furthermore, categories are arranged in hierarchies in order to allow the retention and use of a maximum amount of knowledge with minimal effort (Rosch, 1978). The

multidimensional database approach offers a direct support for categorization and hierarchies via the dimensions and dimension members, grouped by level of granularity, that represent analysis themes. Different users can create different multidimensional models of the same reality according to their own interpretations and particular analytical needs. The different types of dimensions supported by the spatio-temporal multidimensional structure (descriptive, spatial and temporal) can also be related to the idea that humans cognitively store “what”, “where” and “when” knowledge in separate categorical hierarchies that capture differing characteristics for differing purposes (Mennis et al., 2000; Sergent, 1991). Within the multidimensional spatio-temporal structure, spatio-temporal (observational) data of a certain level of detail are collected and incorporated. Then, more general information is derived from the observational data according to the themes and hierarchies defined by the dimensions. This concept is similar to the pyramid framework proposed by Mennis et al. (2000) that decomposes geographic entities into interrelated location-based (where), time-based (when), and object-based (what) information, at the observational data level as well as higher level abstractions.

### 3. An example of SOLAP software

Our research team has been working on SOLAP concepts since 1997. Many SOLAP system prototypes have been developed and implemented over the years for different organizations in diverse application fields (road and maritime transportation, public and environmental health, public security, forestry, agriculture,

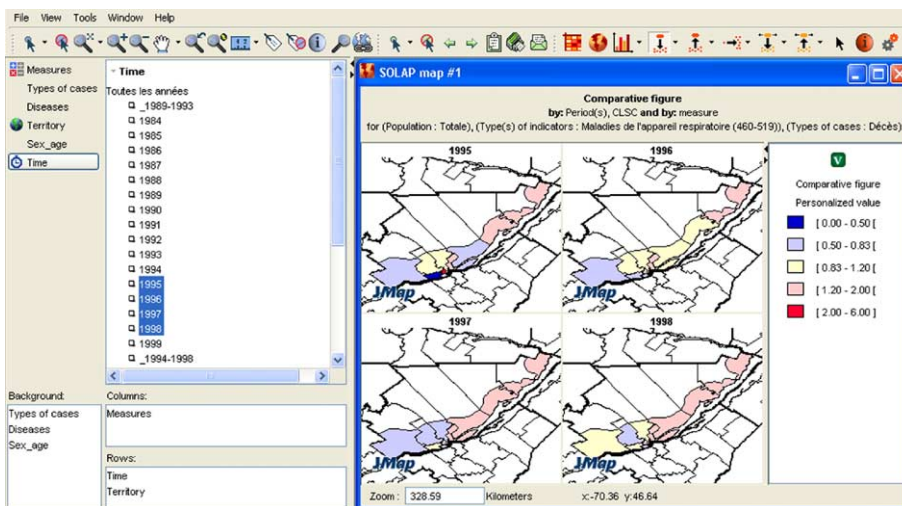


Fig. 3. Temporal multimap showing the respiratory diseases comparative figure of deaths, at the CLSC level, for 1995, 1996, 1997 and 1998.

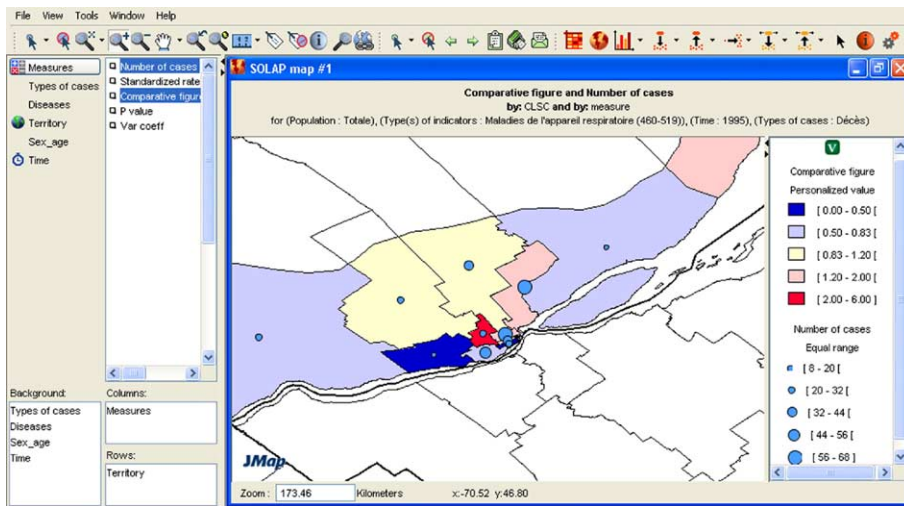


Fig. 4. Thematic map showing the respiratory diseases comparative figure of deaths (polygon color) and the number of deaths (proportional symbols), at the CLSC level, for 1995, and for the region of Quebec. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

archaeology, elite sports, search and rescue) and bench tests have been realized (Gosselin and Bédard, 2002). Following two important GEOIDE research

projects (GEOIDE, 2005) and based on the lessons learned during other prototyping projects, a technology transfer project has been realized in partnership with

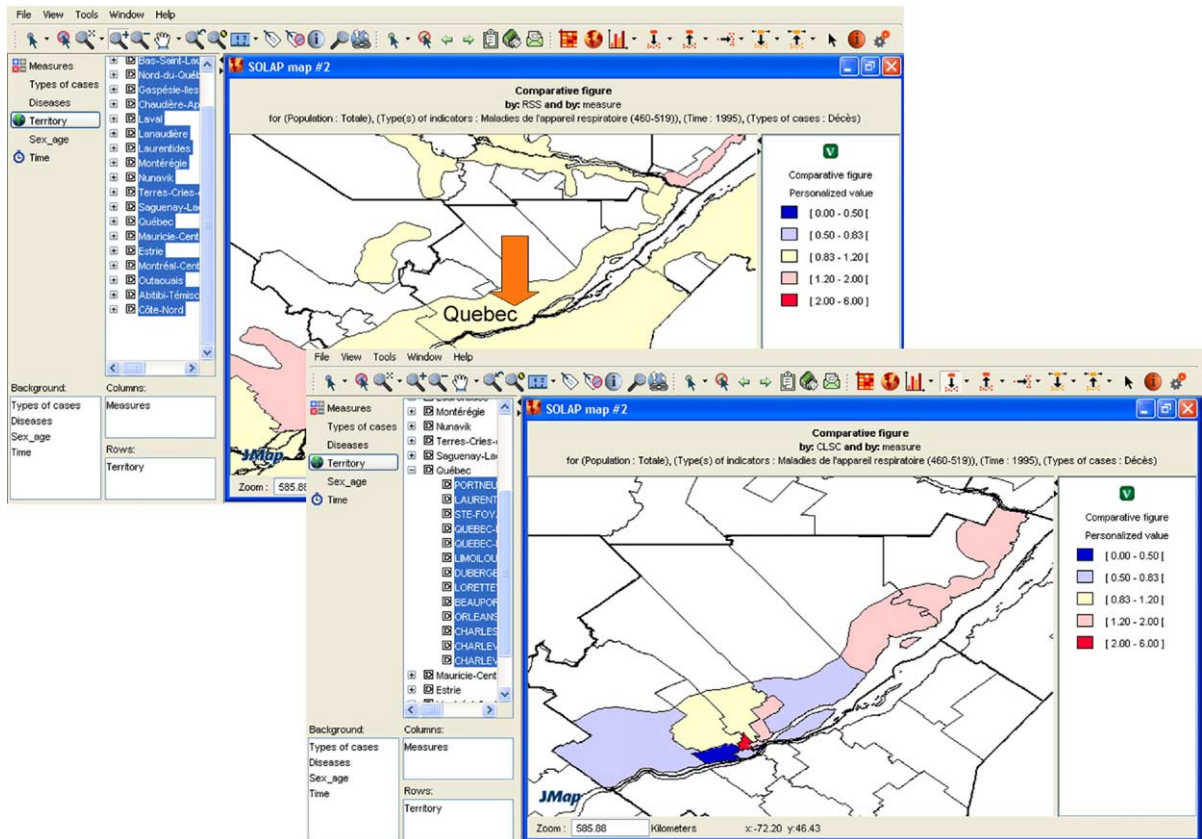


Fig. 5. Spatial drill-down on an RSS polygon (the Quebec RSS), resulting in a map showing the CLSC belonging to the Quebec RSS.



a private company for the development of a generic SOLAP client tool (see [KHEOPS Technologies, 2005](#)). The tool comprises two sections: a database administration module (in the form of wizards) that allows the configuration of the multidimensional spatial database, and a visualization module that allows the interactive display and exploration of data for the end-user. The visualization section offers a series of features that facilitate the spatio-temporal exploration of the data. Some of these features are presented in Sections 3.2–3.6.

The tool is currently based on a ROLAP architecture (without an OLAP server, but bridges to commercial MOLAP servers are planned for the short-term). It sup-

ports the three types of spatial dimensions and the first type of spatial measures discussed in Section 2.2. Pointers are used to link the multidimensional data stored in any commercial DBMS (using a star or snowflake multidimensional schema) to the geometric shapes stored in a spatial format (various spatial format are supported).

### 3.1. An application example

In order to present the different visualization and interactivity features of the developed SOLAP tool, we use an environmental health application example. This particular example uses respiratory diseases data from individual hospitalization data: the Med-Echo

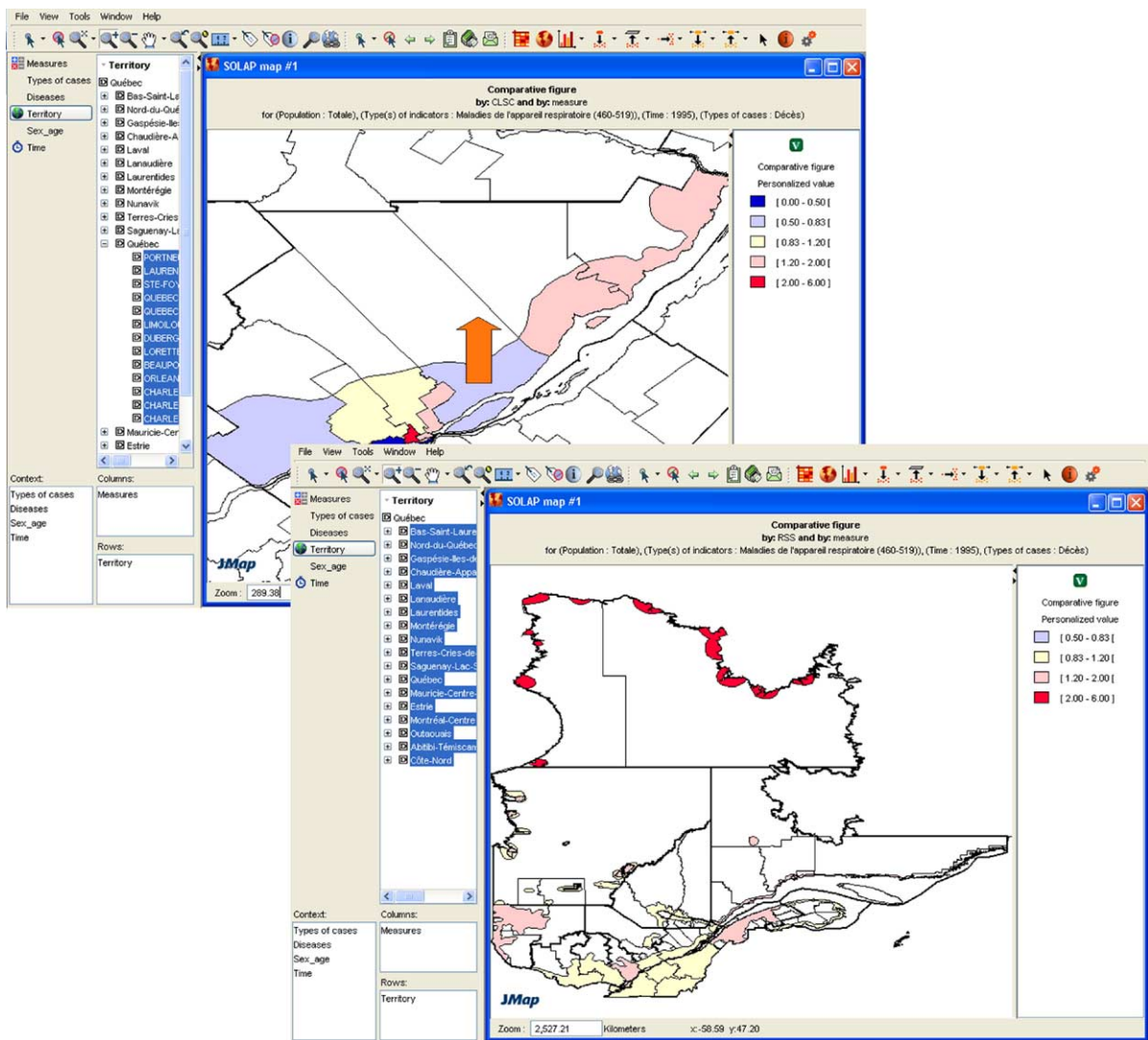


Fig. 6. Spatial roll-up operation on a complete level (the CLSC level), resulting in a map showing all the elements of the next higher level in the dimension, the RSS level.

registry of the Quebec Ministry of Health and Social Services. The temporal coverage is 15 years and the spatial coverage is the province of Quebec, including three levels of aggregation into the spatial hierarchy: the local level (community health centres (CLSC)), the regional level (regional health authorities (RSS)) and the provincial level. For each case (incidence, death or hospitalization), the data collected at the time of the event were: the diagnosis or the death cause (according to the International Classification of Disease, 9th revision), the sex, the age, the event date, the municipal code and the postal code of the individual's principal residence. The postal code has been used to assign the correct CLSC code. The corresponding multidimensional model is then composed of the following dimensions: Disease, Type of case (incidence, death or

hospitalization), Age group, Sex, Territory, and Time; and of the following measures: Number of cases, Standardized rate, Comparative figure, and statistical indicators.

Using this dataset, the SOLAP approach of analysis has been compared to a conventional GIS analysis in (Bédard et al., 2003).

### 3.2. Different types of displays

The developed SOLAP tool currently supports tabular views and 7 types of diagrams (horizontal and vertical bar charts, pie charts, point charts, line charts, area charts and combined (bars and lines) charts). Different types of maps are also possible, according to the types of selections made on the different dimensions: simple

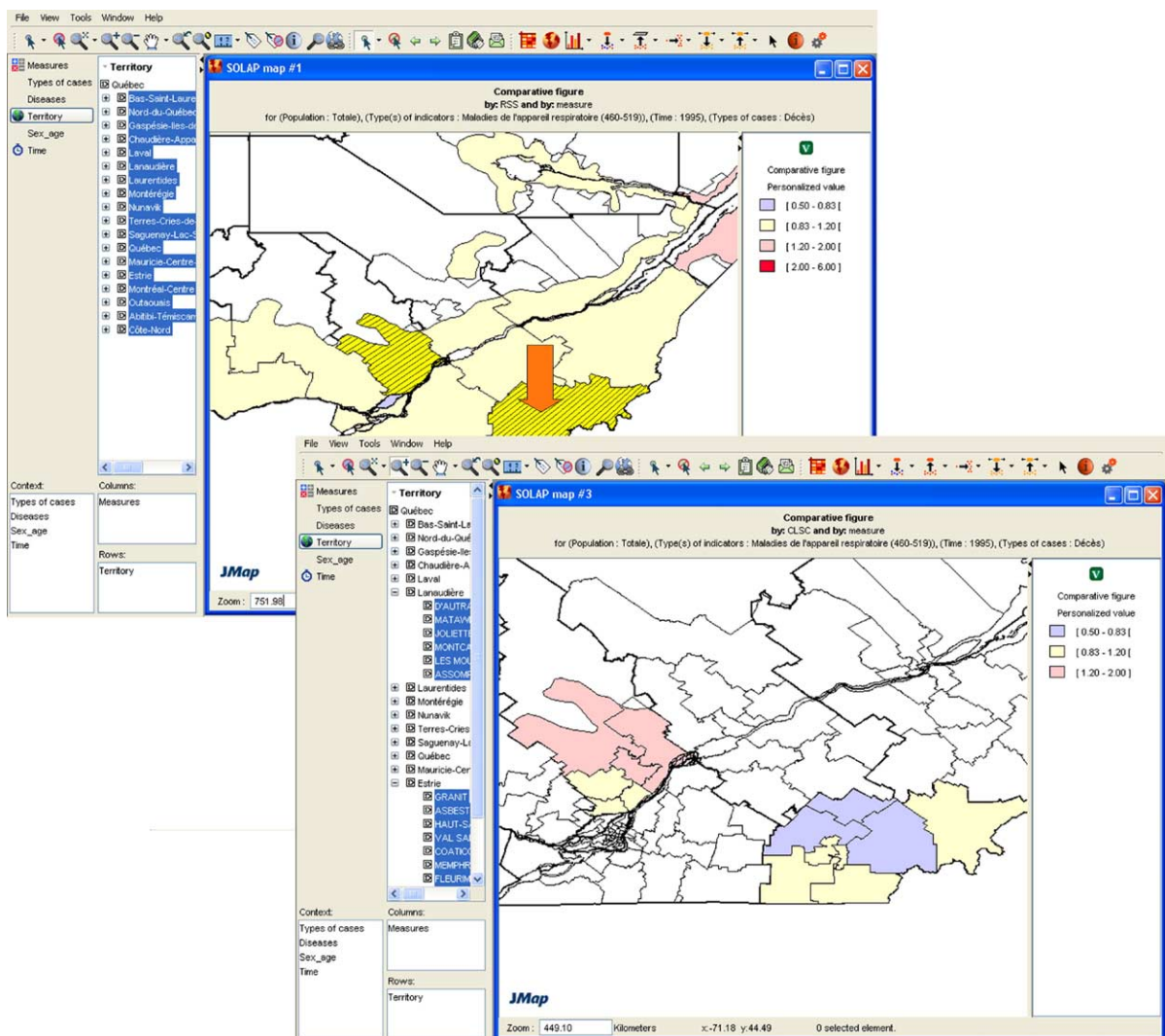


Fig. 7. Spatial drill-down operation on two selected RSS (hatched areas), resulting in a map, showing the CLSC belonging to these RSS.

maps (i.e. single maps showing many geometric elements that reflect a multiple selection on a spatial dimension), multimaps, complex thematic maps (i.e. thematic maps composed of superimposed visual variables (ex. color, pattern, shape of symbols), one per selected measure) and maps with superimposed diagrams (i.e. maps with little histograms or pie charts superimposed to the geometric elements of the map (ex. the different regions) that reflect a number of multiple selections on spatial, temporal or thematic dimensions). Multimaps reflect the concept of small multiples (i.e. a series of maps, for example a map per different year) introduced by (Tuft, 1983), based on the concept of collections (Bertin, 1967). The described SOLAP tool allows for the construction of multimaps using thematic, spatial or temporal dimension members as display multiplication parameters (i.e. a different map for each dimension member). Any number of displays can be opened at the same time and compared. This is a crucial feature to gain alternative perspectives and facilitate data exploration (Gahegan and Brodaric, 2002;

Persson et al., 2005). Fig. 3 presents an example of a temporal multimap showing the respiratory diseases comparative figure of deaths at the CLSC level for 4 different years and for the region of Quebec. Using these maps, it is possible to visualize the evolution of the phenomena under study. Fig. 4 presents an example of a complex thematic map showing the respiratory diseases comparative figure of deaths (polygon color) and the number of deaths (proportional symbols) at the CLSC level, for 1995, and for the region of Quebec. Such maps are typically produced with less than 10 mouse clicks and within 10 s total.

Although the maps produced by the tool are similar to maps produced by current GIS software, the underlying difference is in the way these maps are produced using multidimensional data cubes. The major benefit is a higher level of flexibility and rapidity for interactive data navigation. The objective of this first version of the SOLAP client software was to focus on improved interactivity when navigating in the datasets rather than offering sophisticated visualization tools. Future

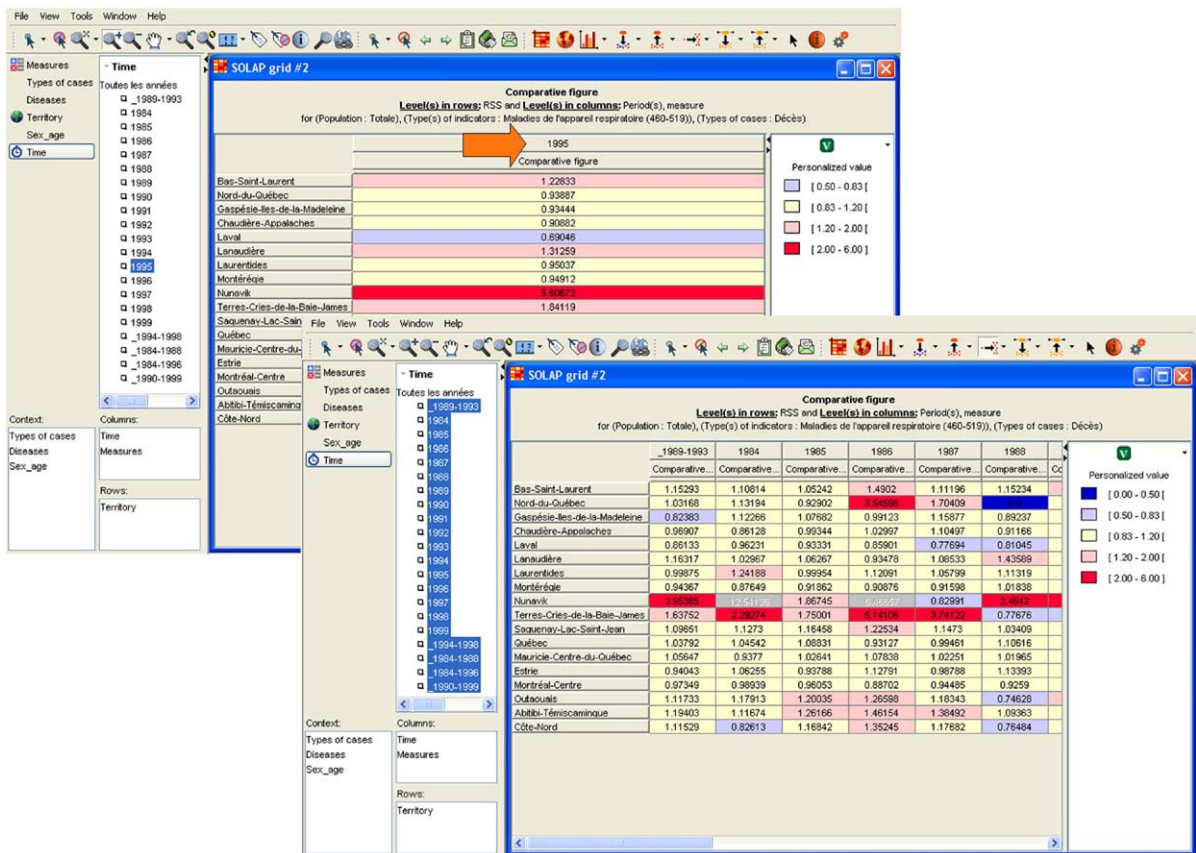


Fig. 8. Temporal drill-across operation on a complete level (the years level of the time dimension) resulting in a table showing all the other elements at the same level of details.

research will consider more advanced visualization tools, such as parallel coordinate plots (Edsall, 2003) for example, for those methods found appropriate for interactive exploration. For example, ongoing research on the use of 3D cartographic visualization has brought extra challenges with regards to the added degrees of freedom for data navigation (Brisebois, 2003).

### 3.3. SOLAP operators

SOLAP operator concepts have been presented in Section 2.2. Different variants of these operators have been implemented: drill-down, roll-up and drill-across on an element clicked with the mouse, on a complete level of detail of a dimension or on a selection of many elements (at the same or at different levels of detail of a dimension). These variants are available for the various

types of operations (thematic, spatial and temporal) and can be used directly on the displayed elements of maps, charts and tables. Two other operators, *open* and *close*, are variants of drill-down and roll-up operators but differ by the fact that they keep the context of the other dimension members. In the next examples presented, the measure shown is the comparative figure of deaths due to respiratory diseases. Fig. 5 presents an example of a spatial drill-down on a RSS polygon (the Quebec RSS) on a map, resulting in a map showing the CLSC belonging to the Quebec RSS. Fig. 6 shows an example of a spatial roll-up operation on a complete level (the CLSC level) resulting in a map showing all the elements of the next level in the dimension, the RSS level. Fig. 7 presents an example of a spatial drill-down operation on selected RSS (hatched areas), resulting in a map showing the CLSC belonging to

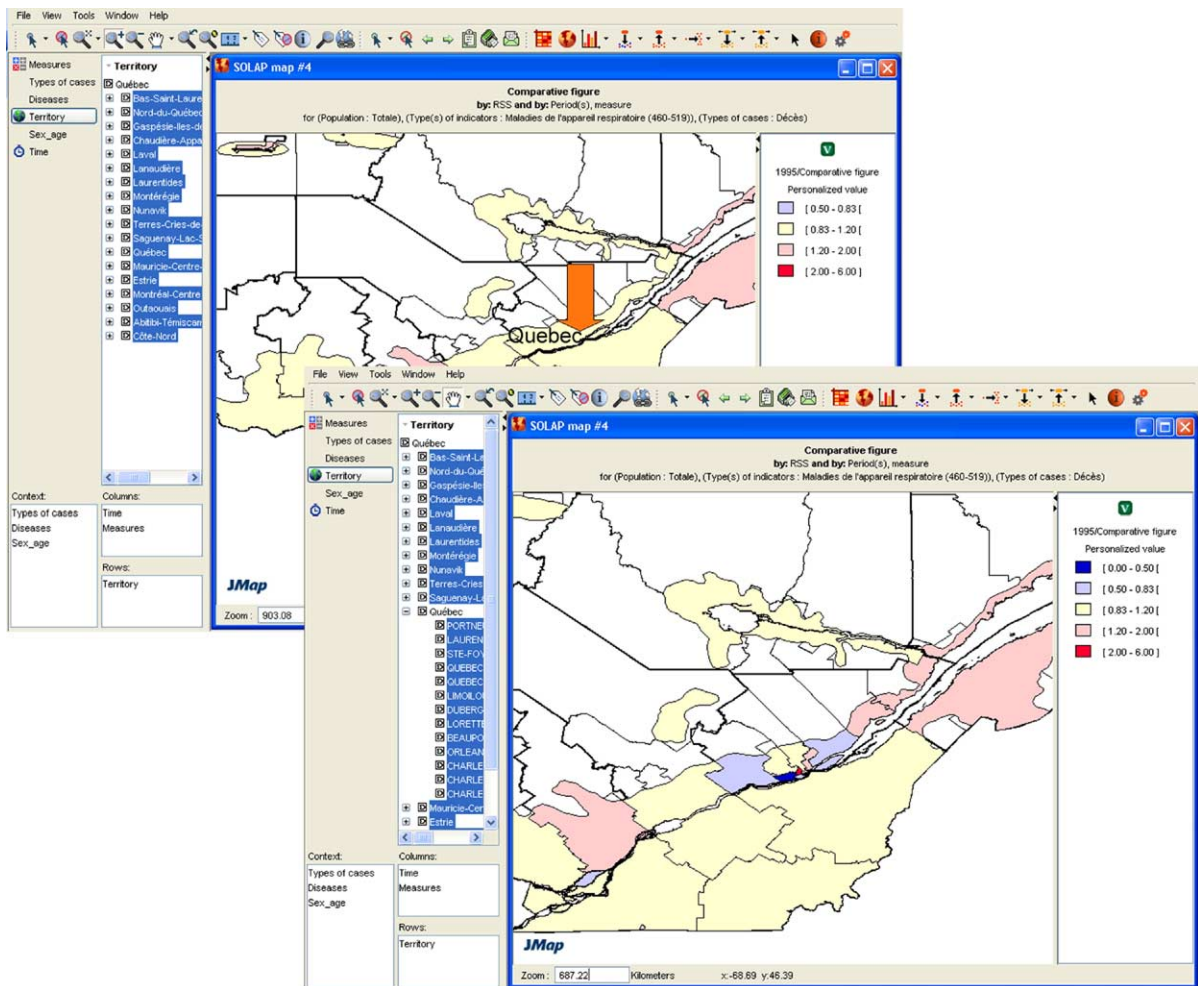


Fig. 9. Open operation on a RSS polygon (the Quebec RSS), resulting in a map showing the CLSC belonging to the Quebec RSS along with the other (un-drilled) RSS.

these RSS. Fig. 8 shows an example of a temporal drill-across operation on a complete level (the years level of the time dimension, in a tabular display) resulting in a table showing all the other elements at the same level of details (here, all the individual years included in the dataset). Fig. 9 presents an example of an open operation on a RSS polygon (the Quebec RSS) on a map, resulting in a map showing the CLSC belonging to the Quebec RSS along with the other (un-drilled) RSS.

### 3.4. Synchronization of the displays and of graphical symbology

The developed SOLAP tool allows two types of synchronization: the synchronization of operations from one display to another, and the synchronization of the graphical symbology used from one display to another. Similar display linking concepts have been proposed by (Andrienko and Andrienko, 2003; Buja et al., 1996). The synchronization of SOLAP operations from one display to another allows the user to visualize the same information, but from a different perspective: the table for the details of the values, the diagrams for rapid comparisons and the maps for the effective visualization of spatial trends or correlations. Temporal comparisons are usually analysed through diagrams, but the SOLAP

tool also offers the possibility to show several maps (what we called *multimaps*), diagrams and tables for different epochs, or for different members of other types of dimensions, and even to browse rapidly through them to simulate dynamic mapping (see Section 3.6). A SOLAP operation executed on one display can be, if this type of synchronization is activated by the user, reflected immediately and automatically in all the other displays in the same *collection* (a collection being a set of linked displays that are synchronized together, any number of collections can be created). The synchronization of the graphical symbology from one display to another cognitively facilitates the identification and the interpretation of the data. Using the same symbology in all the displays, it becomes easier to spot and highlight relevant information. Fig. 10 shows an example of a collection containing a map, a table and a bar chart. An operation on the pie chart will automatically be reflected in the other displays of the collection. However, doing so may lead to potential collisions of graphical symbology rules since theoretically, the same rules do not always apply to maps, pie charts, bar charts, tables, etc. It is necessary to keep a visual homogeneity serving as a link from one display to the other and from one navigation operation to the other. This is even more necessary since the nature of the different types of display used (maps, diagrams and tables) imply that the quantity and nature of information

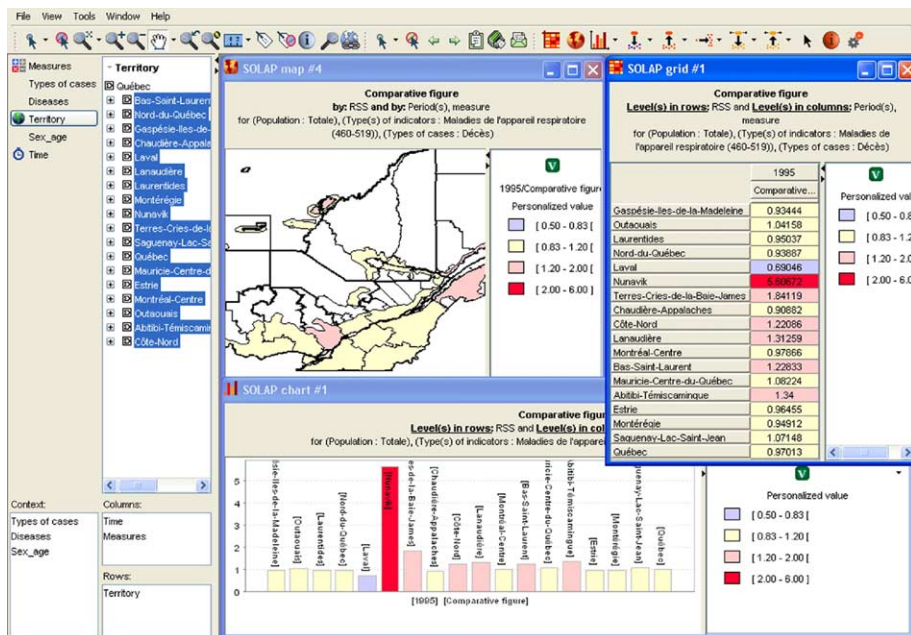


Fig. 10. An example of a collection where the displays and the graphical symbology are synchronized. The map, the table and the bar chart all show the same information: the comparative figure of deaths due to respiratory diseases, for 1995 and for all RSS of the province of Quebec. The same data classification is applied in all the displays and the color visual variable is used.

that can be represented on each display type is different. For example, a table can contain many imbricated axes so the information related to many dimensions can be represented. On a map, a limited number of themes should be represented in order for the map to be visualized as one image and remain readable. There are possible conflicts between the displays and it then becomes necessary to define priorities respecting the way to employ visual variables. A rules manager has been implemented to minimize the potential collisions. The symbology used for representing the different measures in the different displays are established by the administrator of the system (using a flexible symbology manager allowing for different types of data categorization, different types of thematic maps, and the use of various

visual variables). The end-users can also create their own personal symbologies to be applied within their own analysis sessions. Visual variables have been described in details in (Bertin, 1967; Bertin and Barbut, 1999; MacEachren, 1994b, 1995).

The data selection is also synchronized among the different displays of a collection. For example, if a user selects and highlights the polygon corresponding to the Gaspésie region on a map contained in a collection that also contains a table and a pie chart, the Gaspésie region will also be selected and highlighted in the table and in the pie chart. This feature is called brushing. The brushing technique was developed in statistics and is used to relate data points, in a graph or map on a computer screen, with the corresponding entries in the spreadsheet

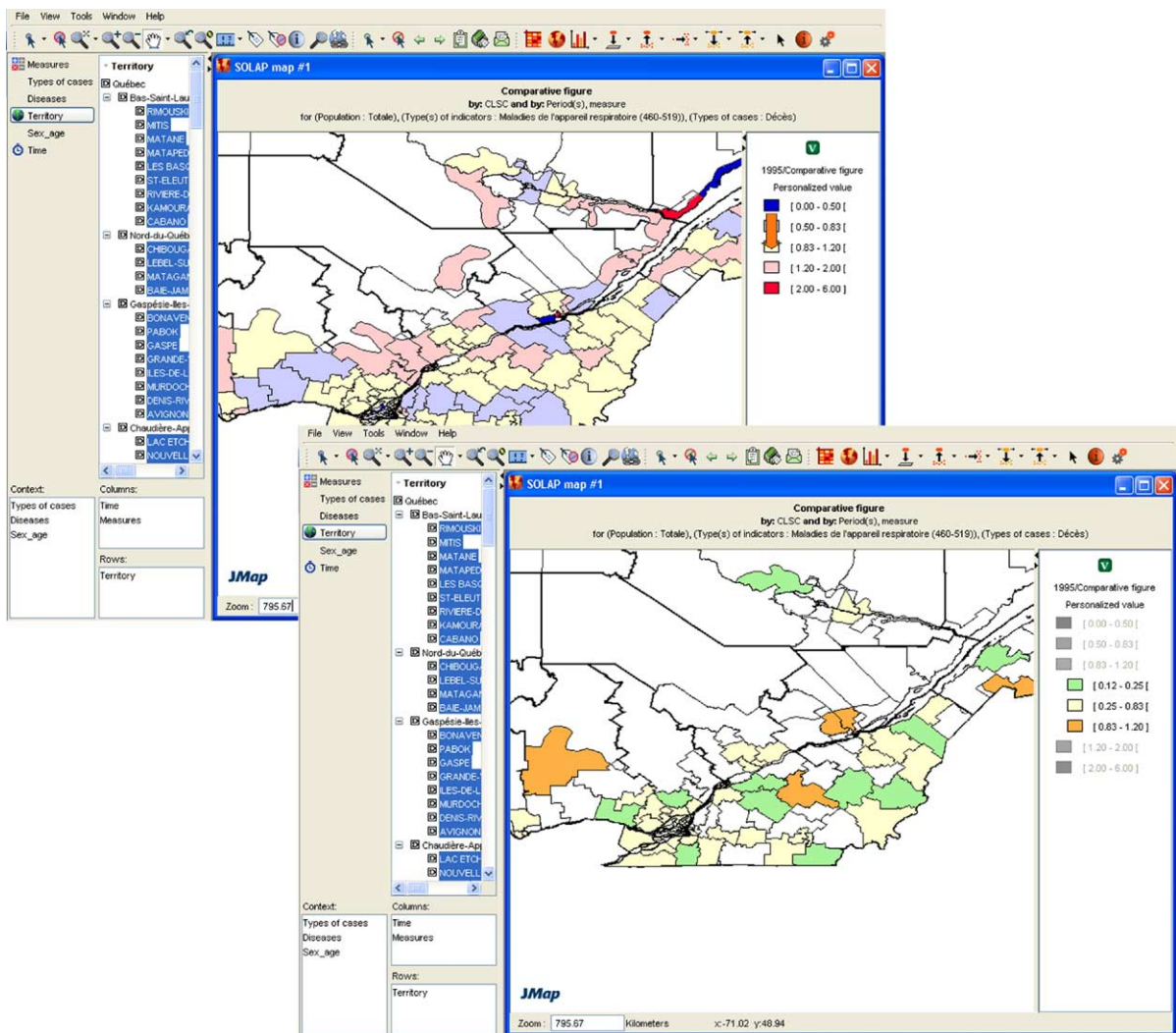


Fig. 11. A drill-down operation on one category of the classification used to group the data to be displayed. The drilled category is refined in a certain number of sub-categories.

from which the graphic was generated (Harris, 1999). When a map display is included in the collection, the term geographic brushing can be used (Monmonier, 1989).

### 3.5. Interactive legend

Pastor (2004) worked at defining elements of interactivity that could be incorporated into the legends of the different displays in a SOLAP interface. The interactive legend can be seen as a graphical view specific to the semantics of the analyzed data. This way, because SOLAP navigation is allowed in all views (or displays) of data, it is also interesting to define navigation capabilities within the legend to remain consistent throughout the user interface. The interactive legend is related to three components: time, space and measures (or “when”, “where” and “what” (Peuquet, 2002)). These three legend components can be represented using different graphical tools. The timeline (see Section 3.6) is an example of tool that can be used for the temporal legend component. With animation, the legend serves not only as an interpretation device but also as a navigation tool (Kraak et al., 1996).

The interactive legend included in the developed SOLAP tool proposes a new type of SOLAP operation: the classification drill operation that is a drill (down or up) applied on the data classification used to represent measures on the different displays. This operation allows for the visualization of different levels of detail of the data classification. Fig. 11 shows an example of this type of drill operation, where one category (i.e. [0.83–1.20]) that groups a certain number of regions on the map is drilled (i.e. exploded in three smaller categories: [0.83–0.95], [0.95–1.08], [1.08–1.20], and the regions originally within this category are redistributed into the sub-categories).

### 3.6. Draggable timeline

The developed SOLAP tool includes a draggable interactive timeline. This timeline allows the user to control the display of the time dimensions and supports the drill-down and roll-up operations (directly on the time cursor). This timeline allows for the display of animated maps and hence, facilitate the visualisation of the evolution of phenomena. The animated maps can depict a trend or a pattern that would not be apparent if the maps were viewed individually (Kraak et al., 1996; Peterson, 1999). Another form of animation is also available with the undo and redo buttons that allow for the visualization of the analysis path followed.

Figs. 3–11 highlight the facility of navigating through the data using mouse clicks in a way that is similar to hyperlinks on web pages (click on what you want). The results of the operations are displayed in a few seconds only, which is rapid compared to the time required to produce the same types of displays using a GIS, for example. Ease-of-use, rapidity, multi-granularity and synchronizable multiviews of information are the keywords that best describe SOLAP when comparing them to the traditional transactional capabilities of GIS.

## 4. Conclusion

SOLAP has been defined as a category of software that allows rapid and easy navigation within spatial databases and that offers many levels of information granularity, many themes, many epochs and many display modes synchronized or not: maps, tables or diagrams. Its multidimensional approach of analysis is more in agreement with the end user’s mental model of the data than the traditional transactional approach typical of GIS. The user interface of such a technology that exploits the multidimensional database paradigm provides unique capabilities to explore data in an intuitive and interactive way. The capability of linking or synchronizing several views of the same information in a context of interactive exploration of data brings new possibilities for benefiting from current research in geovisualization.

This paper reviewed the underlying concepts supporting SOLAP and discussed their use in supporting efficient spatio-temporal exploration of data. Features of a new SOLAP technology have also been presented such as the visualization and manipulation of the geometric component of the spatial data, the database navigation operators, the synchronization of the displays and the graphical symbology, and the interactive legends. Using this technology, the time required to implement a functional system is reduced by an order of magnitude and the access and analysis of spatio-temporal from non-technical users is facilitated.

## Acknowledgements

The authors wish to acknowledge the financial support of Canada NSERC Industrial Research Chair in Geospatial Databases for Decision Support (Hydro-Quebec, National Defence Canada, Natural Resources Canada, Transport Quebec, Intelec Geomatics, KHEOPS Technologies, Holonics, Syntell, DVP, Laval University) and of the GEOIDE Network of Centres of Excellence of

Canada DEC#2 and SOC#1 projects. The authors also thank KHEOPS Technologies for commercializing the resulting software under the commercial name of JMap® Spatial OLAP Extension.

## References

- AltaPlana, 2004. OLAP and OLAP Server Definitions (from OLAP Council). <http://altaplana.com/olap/glossary.html>. Accessed October 12, 2005.
- Andrienko, N., Andrienko, G., 2003. Informed spatial decisions through coordinated views. *Information Visualization* 2 (4), 270–285.
- Bédard, Y., Larrivière, S., Proulx, M.J., Létourneau, F., Caron, P.Y., 1997. Étude de l'état actuel et des besoins de R&D relativement aux architectures et technologies des data warehouses appliquées aux données spatiales. Research Report for National Defence Canada Centre for Research in Geomatics, Laval University.
- Bédard, Y., Merrett, T., Han, J., 2001. Fundamentals of spatial data warehousing for geographic knowledge discovery. In: Miller, H., Han, J. (Eds.), *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis, London, pp. 53–73.
- Bédard, Y., Gosselin, P., Rivest, S., Proulx, M.J., Nadeau, M., Lebel, G., Gagnon, M.F., 2003. Integrating GIS components with knowledge discovery technology for environmental health decision support. *International Journal of Medical Informatics* 70 (1), 79–94.
- Bédard, Y., Proulx, M.J., Rivest, S., 2005. Enrichissement du OLAP pour l'analyse géographique: exemples de réalisation et différentes possibilités technologiques. In: Bentayeb, F., Boussaid, O., Darmont, J., Rabaseda, S. (Eds.), *Entrepôts de Données et Analyse en ligne, RNTI B-1. Cépaduès*, Paris, pp. 1–20.
- Berson, A., Smith, S.J., 1997. *Data Warehousing, Data Mining and OLAP*. McGraw-Hill.
- Bertin, J., 1967. *Sémiologie graphique: les diagrammes, les réseaux, les cartes*. Mouton, Paris.
- Bertin, J., Barbut, M., 1999. *Sémiologie Graphique: les Diagrammes, les Réseaux, les Cartes, 3<sup>e</sup> édition*. Éditions de l'École des Hautes Études en Sciences.
- Brisebois, A., 2003. *Analyse du potentiel d'extension du concept SOLAP pour l'exploration des données spatiales tridimensionnelles*. Unpublished M.Sc. Thesis, Geomatics Sciences Department, Laval University.
- Buja, A., Cook, D., Swayne, D.F., 1996. Interactive High-dimensional Data Visualization. *Journal of Computational and Graphical Statistics* 5 (1), 78–99.
- Caron, P.-Y., 1998. *Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle*. Unpublished M.Sc. Thesis, Geomatics Sciences Department, Laval University.
- Chaudhuri, S., Dayal, U., 1997. An Overview of Data Warehousing and OLAP Technology. *SIGMOD Record* 26 (1), 65–74.
- Codd, E.F., Codd, S.B., Salley, C.T., 1993. *Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate*. Hyperion white papers. [http://dev.hyperion.com/resource\\_library/white\\_papers/providing\\_olap\\_to\\_user\\_analysts.pdf](http://dev.hyperion.com/resource_library/white_papers/providing_olap_to_user_analysts.pdf). Accessed October 12, 2005.
- Date, C.J., 2003. *An Introduction to Database Systems*, 8th edition. Addison-Wesley.
- Edsall, R., 2003. The parallel coordinate plot in action: design and use for geographic visualization. *Computational Statistics & Data Analysis* 43 (4), 605–619.
- Edwards, G., 2001. A virtual test bed in support of cognitively-aware geomatics technologies. *Proceedings of Conference on Spatial Information Theory (COSIT)*, Morro Bay, California, 19–23 September. *Lecture Notes in Computer Science*, vol. 2205. Springer, pp. 140–155.
- Franklin, C., 1992. An introduction to geographic information systems: linking maps to databases. *Database* 15 (2), 13–21.
- Gahegan, M., Brodaric, B., 2002. Computational and visual support for geographic knowledge construction: filling in the gaps between exploration and explanation. *Proceedings of the 10th International Symposium on Spatial Data Handling*, Ottawa, Canada, 9–12 July. *Advances in Spatial Data Handling*, pp. 11–25.
- GEOIDE, 2005. The GEOIDE Network of Centers of Excellence: Geomatics for Informed Decisions. <http://www.geoide.ulaval.ca>. Accessed October 12, 2005.
- Gill, H.S., Rao, P.C., 1996. *The Official Guide to Data Warehousing*. QUE Corporation.
- Gosselin, P., Bédard, Y., 2002. Final Report, GEOIDE Soc#1 Project. The Network of Centres of Excellence in Geomatics (GEOIDE).
- Han, J., Kamber, M., 2001. *Data Mining Concepts and Techniques*. Morgan-Kaufmann.
- Han, J., Stefanovic, N., Koperski, K., 1998. Selective materialization: an efficient method for spatial data cube construction. *Proceedings of the Second Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Melbourne, Australia, 15–17 April, *Lecture Notes in Computer Science* 1394. Springer, pp. 144–158.
- Harris, R.L., 1999. *Information Graphics—A Comprehensive Illustrated Reference*. Oxford University Press, New York.
- Inmon, W.H., 2002. *Building the Data Warehouse*, 3rd edition. Wiley.
- KHEOPS Technologies, 2005. *JMap Spatial OLAP Extension*. <http://www.kheops-tech.com/en/jmap/solap.jsp>. Accessed October 12, 2005.
- Kraak, M.-J., Edsall, R.M., MacEachren, A.M., 1996. Cartographic animation and legends for temporal maps: exploration and/or interaction. *Proceedings of the 7th International Conference on Spatial Data Handling*, The Netherlands, 12–16 August, pp. 17–28.
- LGS Group, 2000. *Analysis of health surveillance business intelligence tools & applications*. Final report for Health Canada.
- MacEachren, A.M., 1994a. Visualization in modern cartography: setting the agenda. In: MacEachren, A.M., Taylor, D.R.F. (Eds.), *Visualization in Modern Cartography*, pp. 1–12.
- MacEachren, A.M., 1994b. *Some Truth with Maps: A Primer on Design and Symbolization*. Association of American Geographers, Washington.
- MacEachren, A.M., 1995. *How Maps Work: Representation Visualization and Design*. Guilford Press.
- MacEachren, A.M., Kraak, M.-J., 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science* 28 (1), 3–12.
- Marchand, P., Brisebois, A., Bédard, Y., Edwards, G., 2004. Implementation and evaluation of a hypercube-based method for spatio-temporal exploration and analysis. *Journal of the International Society of Photogrammetry and Remote Sensing (ISPRS)* 59 (1-2), 6–20.
- Mennis, J.L., Peuquet, D.J., Qian, L., 2000. A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographic Information Science* 14 (6), 501–520.
- Monmonier, M., 1989. Geographical brushing: enhancing exploratory analysis of the scatterplot matrix. *Cartographical Analysis* 21 (1), 81–84.



- Newell, A., 1990. *Unified Theories of Cognition*. Harvard University Press, Cambridge.
- Papadias, D., Tao, Y., Kalnis, P., Zhang, J., 2002. Indexing spatio-temporal data warehouses. *Proceedings of the 18th International Conference on Database Engineering*. San Jose, California, pp. 166–175.
- Pastor, J., 2004. *Conception d'une légende interactive et forable pour le SOLAP*. Unpublished M.Sc. Thesis, Geomatics Sciences Department, Laval University. <http://www.theses.ulaval.ca/>. Accessed October 12, 2005.
- Pendse, N., 2000. *Glossary, The OLAP Report*. <http://www.olapreport.com/glossary.htm>. Accessed October 12, 2005.
- Persson, D., Gartner, G., Buchroithner, M., 2005. Towards a typology of interactivity functions for visual map exploration. *Proceedings of Geo-Hypermedia*, Denver, CO, 4–5 April.
- Peterson, M.P., 1999. Active legends for interactive cartographic animation. *International Journal of Geographic Information Science* 13 (4), 375–383.
- Peuquet, D., 2002. *Representations of Space and Time*. Guilford Press.
- Rafanelli, M., 2003. *Multidimensional Databases: Problems and Solutions*. Idea Group Publishing, London.
- Rivest, S., Bédard, Y., Marchand, P., 2001. Towards better support for spatial decision-making: defining the characteristics of Spatial On-Line Analytical Processing (SOLAP). *Geomatica, the Journal of the Canadian Institute of Geomatics* 55 (4), 539–555.
- Rivest, S., Bédard, Y., Proulx, M.J., Nadeau, M., 2003. SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis. *Proceedings of the ISPRS Joint Workshop on Spatial, Temporal and Multi-Dimensional Data Modelling and Analysis*, Quebec, Canada, 2–3 October.
- Rosch, E., 1978. Principles of categorization. In: Rosch, E., Lloyd, B. (Eds.), *Cognition and Categorization*, pp. 27–77.
- Sergent, J., 1991. Judgements of relative position and distance on representations of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance* 17 (3), 762–780.
- Shekhar, S., Chawla, S., 2002. *Spatial Databases: A Tour*. Prentice-Hall.
- Shekhar, S., Lu, C.T., Tan, X., Chawla, S., Vatsavai, R., 2001. Map cube: a visualization tool for spatial data warehouses. In: Miller, H., Han, J. (Eds.), *Geographic Data Mining and Knowledge Discovery*. Taylor and Francis, London, pp. 74–109.
- Stefanovic, N., 1997. *Design and Implementation of On-Line Analytical Processing (OLAP) of Spatial Data*. Unpublished M.Sc. Thesis, Computer Sciences Department, Simon Fraser University.
- Stefanovic, N., Han, J., Kopersky, K., 2000. Object-based selective materialization for efficient implementation of spatial data cubes. *IEEE Transactions on Knowledge and Data Engineering* 12 (6), 938–958.
- Thomsen, E., 2002. *OLAP Solutions: Building Multidimensional Information Systems*, 2nd edition. John Wiley and Sons.
- Thomsen, E., Spofford, G., Chase, D., 1999. *Microsoft OLAP Solutions*. John Wiley and Sons.
- Tufte, E.R., 1983. *The Visual Display of Quantitative Information*. Graphics Press.