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, Workshop ISPRS, Quebec, Canada, October 2-3, Québec, Canada

SOLAP: A NEW TYPE OF USER INTERFACE TO SUPPORT SPATIO-TEMPORAL MULTIDIMENSIONAL DATA EXPLORATION AND ANALYSIS

S. Rivest, Y. Bédard, M.J. Proulx, M. Nadeau

Centre for Research in Geomatics, 0744 Louis-Jacques-Casault Building,
Laval University (Quebec), Canada, G1K 7P4
yvan.bedard@scg.ulaval.ca

KEY WORDS: Decision-Support, Geographic Knowledge Discovery, Multidimensional Database, On-Line Analytical Processing, Spatial OLAP, User interface

ABSTRACT

It is well known that transactional and analysis systems each require a different database structure. In general, the database structure of transactional systems is optimized for consistency and efficient updates while the database structure of analysis systems is optimized for complex query performance. Non-spatial data are reorganized in data warehouses in order to support analysis and decision-making. In the same way, spatial data need to be stored in *spatial* data warehouses to support spatio-temporal decision-making. However, the actual client tools used to exploit the data warehouse are not well adapted to fully exploit the spatial data warehouse. New client tools are then required to take full advantage of the geometric component of the spatial data. GIS are potential candidates but despite interesting spatio-temporal analysis capabilities, it is recognized that actual GIS systems per se are not optimally designed to be used to support decision applications and that alternative solutions should be used (Bédard et al, 2001). Among them, the Spatial OLAP (SOLAP) tools offer promising possibilities. A SOLAP tool can be defined as "*a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays*" (Bédard, 1997). SOLAP tools form a new family of user interfaces and are meant to be client applications sitting on top of a multi-scale spatial data warehouse. They are based on the multidimensional paradigm. This document presents the concepts of SOLAP, the characteristics of this new type of user interface, and examples related to a few of the many possible application domains. A live demonstration of a SOLAP tool will complete this document.

RÉSUMÉ

Il est bien connu que les systèmes transactionnels et les systèmes d'analyse requièrent une structure de données différente. En général, la structure de données des systèmes transactionnels est optimisée afin de maintenir l'intégrité des données et faciliter la mise à jour, tandis que celle des systèmes d'analyse est optimisée pour la performance de traitement des requêtes complexes. Les données non spatiales doivent être réorganisées sous forme d'entrepôts de données afin de supporter l'analyse et la prise de décision. De la même façon, les données spatiales doivent être stockées à l'intérieur d'entrepôts de données spatiales. Cependant, les outils client actuels utilisés pour exploiter les données des entrepôts ne sont pas pleinement adaptés aux entrepôts de données spatiales. De nouveaux outils client sont requis pour exploiter le plein potentiel de la composante géométrique des données spatiales. Les SIG sont des candidats potentiels, mais malgré leurs capacités d'analyse spatiale poussées, il est reconnu que les SIG seuls, dans leur forme actuelle, ne sont pas bien adaptés pour soutenir des applications d'aide à la décision et que des solutions alternatives doivent être développées (Bédard et al, 2001). Parmi ces solutions, nous retrouvons les outils OLAP spatiaux (SOLAP), qui offrent des possibilités prometteuses. Un outil SOLAP peut être défini comme "*une plate-forme visuelle spécialement conçue pour supporter l'analyse et l'exploration spatio-temporelles rapides et faciles des données multidimensionnelles composées de plusieurs niveaux d'agrégation à l'aide d'affichages cartographiques aussi bien qu'à l'aide de tableaux et diagrammes statistiques*" (Bédard, 1997). Les outils SOLAP forment une nouvelle famille d'interfaces à l'utilisateur et sont conçus comme des applications client d'entrepôts de données spatiales multiéchelles. Ils sont basés sur le paradigme multidimensionnel. Ce document présente les concepts associés aux outils SOLAP, les caractéristiques de ce nouveau type d'interface à l'utilisateur, ainsi que des exemples de quelques-uns des domaines d'application possibles. Une démonstration de l'utilisation d'un outil SOLAP complétera ce document.

1. INTRODUCTION

The advent of data warehouses and their popularity changed the way people work with data. The transactional (On-Line Transaction Processing or OLTP) systems, often called legacy systems in the data warehouse terminology, were designed to support the everyday transactions of the organizations. They were not efficient in supporting the analysis processes required for decision-making, and this, because of their intrinsic database structure. New systems, based on new concepts, were then specifically designed to facilitate the analysis processes. These systems, in which the data warehouse is usually a central component, also comprise different types of client applications allowing users to analyze and explore their data.

It has been estimated that about 80 percent of all data stored in corporate databases are spatial data (Franklin, 1992). In order to store the geometric component of the spatial data inside data warehouses, organizations build *spatial* data warehouses. However, the common data warehouse clients: query and report builders, On-Line Analytical Processing (OLAP) tools and data mining applications, are not optimized to explore and analyze the spatial data. They can be used, but without the capability to manipulate the geometric component of the data, they cannot provide a thorough analysis.

Geographic Information Systems (GIS) are potential candidates to serve as client applications to spatial data warehouses. However, despite interesting spatio-temporal analysis capabilities, it is recognized that actual GIS per se, with often complex query interfaces and slow response times, are not adequate for decision-support applications and that alternative tools should be used (Bédard et al, 2001). Among the possible solutions, the coupling of spatial and non-spatial technologies, GIS and OLAP for instance, may appear to be a good option. This coupling paves the way to the emergence of a new family of OLAP tools, better adapted for spatial, or spatio-temporal decisional analysis: spatial OLAP, or SOLAP, applications.

Sometimes qualified as "keyboardless-GIS", but being much more than that in terms of navigation within multidimensional datasets, SOLAP tools possess an intuitive user interface allowing non-technical users to easily access, visualize and analyze their data.

Today, SOLAP has become a viable solution to explore spatio-temporal data and the most recent research aims at improving the underlying fundamental concepts as well as the capabilities of emerging commercial offerings.

This paper briefly reviews the general concepts of OLAP tools and then introduces and discusses the SOLAP tools and their general characteristics. The following section discusses the characteristics related to the particular user interface of SOLAP tools. The next section presents the actual research topics related to SOLAP. Finally, the last section presents examples of the use of SOLAP tools in different application domains.

2. OLAP AND SOLAP CONCEPTS

2.1 OLAP concepts

OLAP has first been defined 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). Other definitions have been proposed more recently, including "A software category intended for the rapid exploration and analysis of data based on a multidimensional approach with several aggregation levels" (Caron, 1998).

OLAP technology is based on the multidimensional database approach. This approach introduces new concepts, which include dimensions, members, granularity, measures, facts and data cubes. The *dimensions* represent the analysis themes, or the analysis axis (e.g. "time", "products", "sales territory"). A dimension contains *members* (e.g. "1998", "shirt", "Quebec region") that are organized hierarchically into levels of *granularity*, or levels of details (e.g. "city", "region", "province", "country" for a "sales territory" dimension). The members at one level (e.g. "city") can be grouped, or aggregated, to form the members of the next higher level (e.g. "region") inside a dimension. Different types of dimension can be defined: temporal dimensions, spatial dimensions (non-cartographic in the case of a conventional OLAP tool) and descriptive (or thematic) dimensions. The *measures* (e.g. sales in dollars, profits) are numerical values analyzed against the different dimensions. A measure can be considered as being the dependent variable while dimensions are the independent variables (e.g. the measure "sales in dollars" depends on the members of the "time", "product" and "sales territory" dimensions). The different combinations of dimension members and measures values represent *facts* (e.g. "the sales of shirts, in 1998, for the Quebec region are 356 763$"). A set of measures aggregated according to a set of dimensions is called a *data cube* (or *hypercube* if there are more than 3 dimensions). Inside a data cube, the possible aggregations of measures on all the possible combinations of dimension members (the facts) can be pre-computed to increase query performance. Figure 1 presents the multidimensional database concepts.

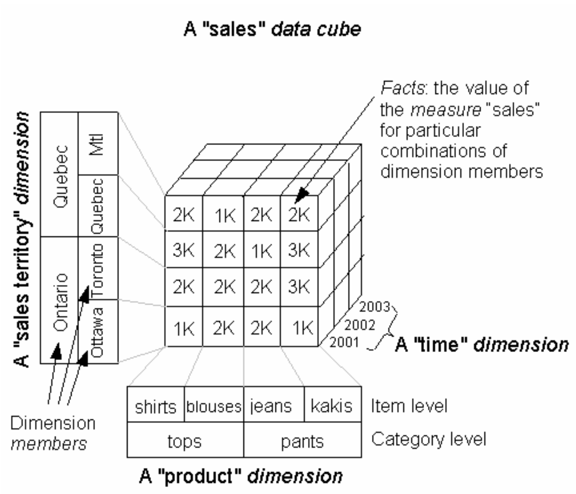

Figure 1. A data cube showing the multidimensional database concepts

The general architecture of an OLAP application is usually composed of three elements: the database, that supports the multidimensional data structure described earlier; the OLAP server, that manages the database and carries out the different calculations; and the OLAP client that allows the end user to explore and analyze the data using different visualization methods and adapted operators.

Depending on the technology 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), (Archer Decision Sciences, 1995).

2.2 SOLAP Concepts

With the development of spatial data warehouses concepts, it became evident that the common client tools used to exploit non-spatial data warehouses were not sufficient to fully analyze the geometric spatial component of the spatial data, this geometric component representing 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 tools into a new type of application. This combination gave birth to Spatial OLAP or SOLAP applications.

SOLAP can be defined as "a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays" (Bédard, 1997). SOLAP are a type of client applications sitting on top of multi-scale spatial data warehouses (Bédard et al. 2001). However the non-expert can also see them as a new type of user interface for multi-scale GIS applications and web mapping.

SOLAP tools support the multidimensional database structure described in section 2.1. However, with the spatial data manipulation capabilities included, three types of spatial dimensions can be used (Han et al, 1998): the non-geometric spatial dimensions, the geometric spatial dimensions and the mixed spatial dimensions. In the first type of spatial dimension, the spatial reference uses nominal data only (e.g. place names) as no geometry or cartographic representation is associated to the dimension members. This type of spatial dimension is the one currently used within conventional OLAP tools. The two other types of spatial dimensions include geometric shapes spatially referenced on a map to allow their dimension members to be visualized and queried cartographically. These geometries exist for all the levels in the case of geometric spatial dimensions, and for some of the levels in the case of mixed spatial dimensions. Figure 2 presents the three types of spatial dimensions.

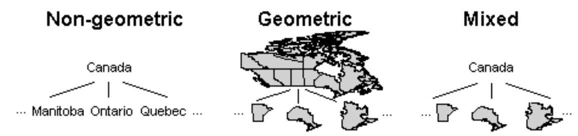

Figure 2. The three types of spatial dimensions supported by SOLAP tools

Also because of their spatial data manipulation capabilities, SOLAP tools can support different types of spatial measures (Rivest et al, 2001), in addition to the numerical measures supported in conventional OLAP tools. A first type consists of a geometric shape or set of shapes being obtained by the combination of multiple geometric spatial dimensions. It consists of a set of coordinates, which requires a geometric operation, such as spatial merge or spatial intersection, to be computed. An example of a geometric spatial measure would be the set of intersected polygons resulting from the combination of the polygonal members of two geometric spatial dimensions such as a political boundaries dimension and a watersheds dimension. 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". Currently, not all these types of spatial measures are fully implemented and it may be necessary to use pointers (stored within the multidimensional data structure) to the geometric shapes stored in another structure or software. Actual research topics aim at determining the feasibility of some of these types of measures.

Like 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. The SOLAP database stores the dimension data as well as the measure values. Because the dimension members and the measure values can comprise a geometric component, the database management system used must support this type of data. The SOLAP server has to handle the spatio-temporal multidimensional database and the numerical and spatial calculations necessary to compute the measure values associated to all the possible combinations of dimension members. Currently, no such server is available on the market. It must then be implemented using a custom combination of existing technologies. The SOLAP client will be discussed in more details in the next section.

3. SOLAP: A NEW TYPE OF USER INTERFACE

3.1 Components of the interface

The interface of a SOLAP tool is composed of two main parts: the visualization space and the navigation panel. The visualization space allows the user to see the desired information in the form of one to many map(s), statistical diagram(s) or table(s) with the possibility to change the displays in order to see the information in the most meaningful way and with the possibility to synchronize all the displays. All the displays are dynamic in the sense that the user can use different operators (that will be described in 3.3) to "navigate" inside the dataset through the displayed elements (e.g. a polygon on the map, a bar in a bar chart, a cell in a table). In the context of information exploration, maps and graphics do more than make data visible; they are active instruments in the end-users thinking process (MacEachren and Kraak, 2001). All the possible views (displayed as maps, diagrams or tables) do not have to be prepared in advance, but are dynamically managed by the SOLAP server.

The navigation panel allows the user to select one or more members of each dimension, and also the measures, to be viewed in the visualization space. All the possible combinations of dimension members and measures are available (unless a filter is used) and are presented to the user through lists or tree views. The user clicks on the elements in the lists and the displays are automatically updated to reflect the new selection (there is no need to use the keyboard).

3.2 Intuitive interface

The multidimensional analysis approach is in agreement with the end user's mental model of the data (Codd et al, 1993). Based on this approach, the interface of an application based on this paradigm, such as SOLAP, is very intuitive and the user can design the analysis by clicking on the data itself (Yougworth, 1995).

The two key elements that best describe SOLAP tools are ease of use and fast response times. First, the ease of use comes from 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, 2003). With a SOLAP tool, the analyst focuses on the results of the analysis rather than on the procedure required by the tool to perform the analysis process. Second, it is rapid because data are pre-aggregated, computation time is then reduced and very fast answers to complex queries are possible. This allows the user to maintain his train of thought, his attention not being distracted by slow response times.

3.3 New operators

In the SOLAP interface, new operators are defined in order to take advantage of the multidimensional data structure and of the different levels of details of the data. The main operators are drill-down, roll-up (or drill-up), drill-across and swap.

A drill-down operation allows the end-user to navigate from a general level to a more detailed level inside a dimension (e.g. from visualizing the province data to visualizing the region data). A roll-up operation allows, conversely, the user to navigate from a detailed level to a more general level inside a dimension (e.g. from visualizing the region data to visualizing the province data). A drill-across allows to view different information but at the same level of detail (e.g. from the Montreal region to the Quebec region, or from a sales measure to a profits measure). A swap operation allows to interchange two dimensions in order to view the same information in a different way or in order to view different information. These operations are available, in a SOLAP application, in the different types of displays (maps, statistical diagrams or tables). When defined to manipulate the data on maps, the drill operators can be named *spatial* drill-down, *spatial* roll-up and *spatial* drill-across. They allow to navigate from one geometric level of details to another inside a geometric or a mixed spatial dimension, while keeping the same level of thematic granularity.

Thematic operations, that is drill-down, roll-up and drill-across operations in the thematic data while keeping the same level of spatial granularity, are also defined. They can be available from the legend of the displays.

4. ACTUAL RESEARCH TOPICS

The Center for Research in Geomatics of Laval University currently works at defining new concepts to improve SOLAP tools. Work items include the definition of spatio-temporal topological operators dimensions, the concepts supporting a 3D SOLAP tool and the possible improvements of the general user interface.

(Marchand et al, 2003) have proposed a method that implements a Spatio-Temporal Topological Operators Dimension (ST^2OD) into multidimensional databases. The hierarchy of this dimension is structured according to multiple levels of granularity of topological relationships. Used in conjunction with other dimensions, the ST^2OD produces cross-dimensional views that satisfy specific spatial and temporal topological constraints. End-users can navigate within the ST^2OD hierarchy from highly aggregated spatio-temporal topological operators (e.g. “same time, same place”) to finer granularities (e.g. “started by, interior intersection”) down to the finest granularities (e.g. “1--00-102; --1--0102” based on the ISO/TC211 (ISO TC211, 2003) e-Relate operator (Egenhofer and Franzosa, 1991; Egenhofer et al, 1994). This approach can be replicated with metric operators (e.g. distance, area) as well as other more complex operators (e.g. visibility and slope operators) and improves the processing time of the different types of operations.

Another actual research topic consists of determining the concepts supporting a 3D version of a SOLAP tool. Brisebois (Brisebois, 2003) first worked at defining the types of environments and reference systems in which 3D SOLAP tools can operate. Then, characteristics and SOLAP operators particular to the addition of the third dimension were studied. The concepts were tested with applications in forestry and archaeology.

Some team members are also working on different components of the SOLAP interface, the timeline and the legend for instance, to improve their usability and their functionalities in order for the interface to be even more intuitive. One topic concerns the design of an interactive legend in order for the user to navigate within the data through the different displayed components of this legend.

5. A FEW APPLICATION DOMAINS

SOLAP tools can be used in a multitude of application domains. Within our different research projects, our team developed prototype applications in different fields.

– Road transportation: to help better distribute the budgets for road maintenance, according to different parameters (dimensions): time, road conditions, road classifications, traffic flows, etc. and to simulate road conditions degradation. Figure 3 shows this application.

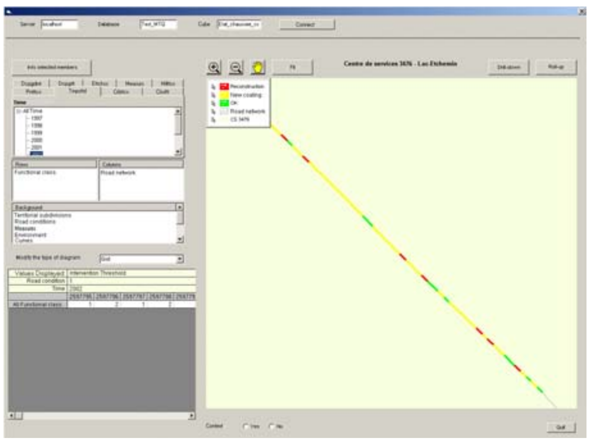

Figure 3. A road transportation application: visualization of an intervention threshold for the segments of a particular road.

– Marine transportation: to analyse the origin and destination of goods according to time, category of goods, itinerary, etc. Figure 4 illustrates this application.

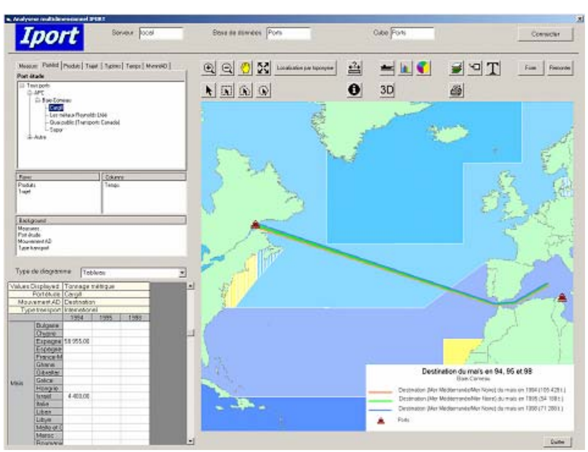

Figure 4. A marine transportation application: visualization of the destination of corn from the port of Baie-Comeau.

– Public security: to support a better and easier crime analysis according to numerous parameters: type of crime, age and gender of the victim, age and gender of the presumed guilty, type of crime scene, etc. Figure 6 illustrates this example.

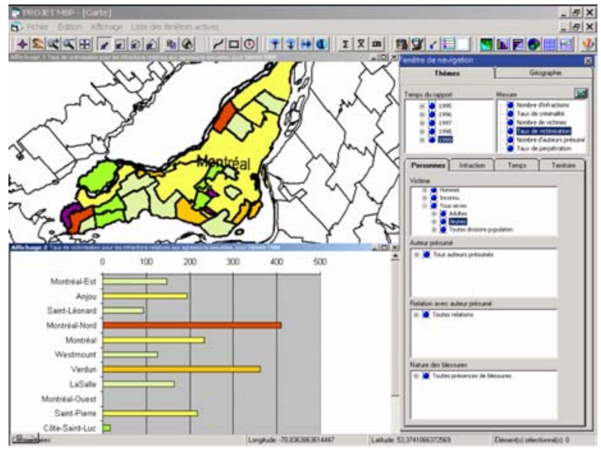

Figure 5. A public security application: visualization of the victim rate for sexual abuse crimes.

– Environmental health: to explore possible relations between health and environmental phenomena, for example the respiratory diseases and the air quality, and to rapidly find or discard hypotheses. Figure 5 presents such an example.

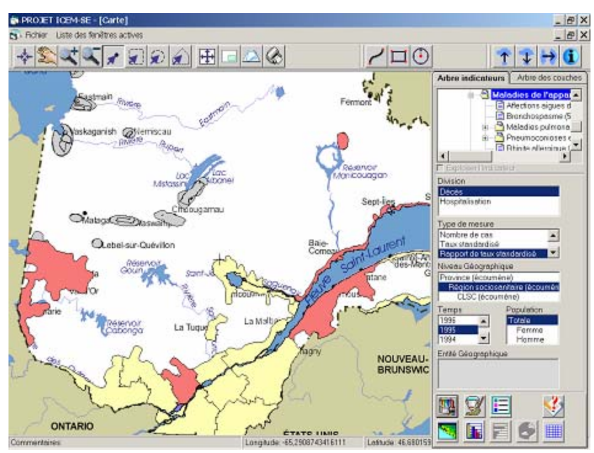

Figure 6. An environmental health application: visualization of the comparative death figure from respiratory diseases.

This application will be presented in details in the demonstration of a SOLAP tool.

6. CONCLUSION

This paper presented a new category of OLAP tools that are designed for effective spatio-temporal exploration and analysis: Spatial OLAP (SOLAP) tools. SOLAP is defined as “a visual platform built especially to support rapid and easy spatio-temporal analysis and exploration of data following a multidimensional approach comprised of aggregation levels available in cartographic displays as well as in tabular and diagram displays”. The general concepts of SOLAP have been presented and also the actual research topics in this field. A few application examples were also included. However, SOLAP tools can be employed beneficially in numerous other domains.

7. REFERENCES

Archer Decision Sciences. 1995. Star Schema 101: A pictorial Introduction to Dimensional Modeling. http://www.archer-decision.com/knowpre.htm.

Bédard, Y. 1997, Novembre. Spatial OLAP, Vidéoconférene. *2e Forum annuel sur la R-D, Géomatique VI: Un monde accessible*, Montréal, Canada.

Bédard, Y., Merrett, T., Han, J. 2001. Fundamentals of spatial data warehousing for geographic knowledge discovery. In H. Miller & J. Han (Eds.), *Geographic Data Mining and Knowledge Discovery*. London: Taylor & Francis.

Brisebois, A. 2003. Analyse du potentiel d’extension du concept SOLAP pour l’exploration des données spatiales tridimensionnelles. Mémoire de M.Sc., Département des sciences géomatiques, Faculté de foresterie et géomatique, Université Laval, 120 p.

Caron, P.-Y. 1998. Étude du potentiel de OLAP pour supporter l'analyse spatio-temporelle. Mémoire de M.Sc., Département des sciences géomatiques, Faculté de foresterie et géomatique, Université Laval, 132 p.

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, 20 p.

Egenhofer, M.F., Franzosa, R. 1991. *Point Set* Topological Spatial Relations, *International Journal of Geographical Information Systems*, vol 5, no 2, pp. 161-174.

Egenhofer, M.J., Clementini, E., Di Felice, P. 1994. Topological relations between regions with holes, *International Journal of Geographical Information Systems*, vol 8, no 2, pp 129-142.

Franklin, C. 1992. An Introduction to Geographic Information Systems: Linking Maps to Databases. *Database*, April, pp. 13-21.

Gill, H.S., Rao, P.C. 1996. *The official Guide to Data Warehousing,* QUE Corporation, 382 p.

Han, J., Stefanovic, N., Koperski, K. 1998. Selective materialization: An Efficient Method for Spatial Data Cube Construction. *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD'98)*, Melbourne, Australia.

ISO TC211. 2003. 19107 – Geographic Information – Spatial Schema. International Standard.

MacEachren, A.M., Kraak, M.-J. 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science*. 28(1).

Marchand, P., Brisebois, A., Bédard, Y. et Edwards G. 2003. 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)*. Accepted for publication.

Pendse, N. 2000. Glossary, The OLAP report. http://www.olapreport.com/glossary.htm.

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), pp. 539-555.

Yougworth, P. 1995. OLAP Spells Success For Users and Developers. *Data Based Advisor*, December, pp. 38-49.

8. ACKNOWLEGDEMENT

This research has been conducted with the financial support of the GEOIDE Network of Centres of Excellence DEC#2 (Designing the technological foundations of geospatial decision-making with the World Wide Web) and SOC#1 (Cartographic interface for the multidimensional exploration of environmental health indicators) projects and the Natural Sciences and Engineering Research Council of Canada individual research grant program.