Bédard, Y., Rivest, S. & Proulx, M.-J., 2005, Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective, Data Warehouses and OLAP : Concepts, Architectures and Solutions, EDA Group, chapitre de livre soumis pour publication en 2006

Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective.

Yvan Bédard*, Sonia Rivest **, Marie-Josée Proulx ***

Canada NSERC Industrial Research Chair in Geospatial Databases for Decision-Support
Centre for Research in Geomatics,
0744 Casault Hall,
Geomatics Sciences Department
Faculty of Forestry and Geomatics
Laval University, Quebec City, Canada,
G1K 7P4
Phone: 418-656-2131 ext. 3694
Fax: 418-656-7411

*Yvan.Bedard@scg.ulaval.ca
http://sirs.scg.ulaval.ca/YvanBedard
** Sonia.Rivest@scg.ulaval.ca
*** Marie-Josee.Proulx@scg.ulaval.ca

Spatial On-Line Analytical Processing (SOLAP): Concepts, Architectures and Solutions from a Geomatics Engineering Perspective.

ABSTRACT

It is recognized that 80% of data have a spatial component (ex. street address, place name, geographic coordinates, map coordinates). Having the possibilities to display data on maps, to compare maps of different phenomena or epochs, and to combine maps with tables and statistical charts allows one to get more insights into spatial datasets. Furthermore, performing fast spatio-temporal analysis, interactively exploring the data by drilling on maps similarly to drilling on tables and charts, and easily synchronizing such operations among these views is nowadays required by more and more users. This can be done by combining Geographical Information Systems (GIS) with On-Line Analytical Processing (OLAP), paving the way to "SOLAP" (Spatial OLAP). The present chapter focuses on the spatial characteristics of SOLAP from a geomatics engineering point of view: concepts, architectures, tools and remaining challenges.

Keywords: Decision Support Systems (DSS), Geographic Information Systems, Knowledge Discovery

INTRODUCTION

It is recognized that up to 80% of corporate data have spatial components such as street addresses, place names, geographic coordinates or map coordinates. This fact, estimated by Franklin (1992), is only starting to show its potential for the masses with recent commercial advances such as Google Maps and Google Earth. However, the true power of maps typically remains underused for geographic knowledge discovery unless one combines a Geographic Information System (GIS) or a universal server (ex. Oracle Spatial) to OLAP technology.

The power of maps

Map data are the raw material to produce the geographic information that leads to knowledge about the position, extent and distribution of phenomena over our territories. Such phenomena are counted by thousands and include insect territorial expansions, environment-health correlations, land-use evolution, 911 vehicle tracking and watershed analysis to name a few. Visualizing geographic phenomena on maps facilitates the extraction of insights that help to understand these phenomena. Such insights include spatial characteristics (position, shape, size, orientation, etc.), spatial relationships (adjacency, connectivity, inclusion, proximity, exclusion, overlay, etc.) and spatial distribution (concentrated, scattered, grouped, regular, etc.). When we visualize a map displaying different regions, we can compare. When we visualize different maps for a same region, we can discover correlations between phenomena. When we visualize the map of a region for different epochs, we can see the evolution of the phenomena. When we use maps, we often get a better understanding of the structures and relationships contained within spatial datasets than using simple tables and charts. When we combine maps with tables and statistical charts, we can relate these to make new discoveries. Maps are natural aids to the knowledge discovery process. In the context of spatial data exploration, maps do more than just make the data visible, they are active instruments to support the end-users thinking process. Using maps for geographic knowledge discovery requires less abstraction efforts for users, which in turn, increases their efficiency. Maps show information that would not be visible from non-spatial data for those phenomena having a spatial distribution that doesn’t correspond to predefined boundaries (ex. administrative boundaries). Numerous studies in cognitive sciences have shown the superiority of images over numbers and words to stimulate understanding and memory (Buzan & Buzan, 2003; Fortin & Rousseau, 1989; Standing, 1973), leading to a more efficient knowledge discovery process (more alert brain, better visual rhythm, more global perception).

Marrying OLAP with GIS

Geographic Information Systems (GIS) are very good at achieving the goal they have been developed for, that is gathering, storing, manipulating and displaying spatial data (see Longley, Goodchild, Maguire, & Rhind, 2001). However, they are transaction-oriented systems and do not address summarized information, cross-referenced information, interactive exploration of data, etc. Furthermore, they are not suited for temporal data, they are very slow to aggregate data, they hardly deal with multiple levels of data granularity, and their user interface is too complex for most users. Similarly to Database Management Systems (DBMS), GIS alone cannot fill the “analysis gap” between spatial data and geographic knowledge discovery. For the typical process of geographic knowledge discovery, the response time of GIS goes well beyond Newell’s cognitive band of 10 seconds (Newell, 1990) and therefore is not satisfactory to support this process interactively. Nevertheless, it has been demonstrated that it is possible to achieve such performance by marrying GIS with OLAP and building efficient spatial datacubes (Marchand, 2004). In addition, several projects marrying GIS and OLAP have shown the superiority of this combination for spatial interactive data exploration over stand-alone GIS or OLAP. On the other hand, simply coupling OLAP tools with GIS is not enough. Even though OLAP are well-suited for knowledge discovery, they are not adapted for the analysis of spatial data. In fact, spatial data is typically treated like other descriptive data and spatial analysis is limited to predefined nominal locations (ex. names of countries, states, regions, cities).. Support for spatio-temporal analyses is seriously limited (no spatial visualization, practically no spatial analysis, no map-based exploration of data, etc.). Standard extraction, transformation and load (ETL) processes cannot deal with most aspects of spatial data (nor the GIS). Coupling GIS and OLAP is not sufficient to solve these issues; many hidden challenges must be overcome, resulting into important development efforts before obtaining an efficient solution.

Towards SOLAP

The LGS Group has introduced three approaches to integrate GIS and OLAP: OLAP-centric, GIS-centric and hybrid (LGS Group, 2000). These approaches require important development efforts and correspond to the solutions offered by major GIS, DBMS and OLAP companies. Bédard, Proulx and Rivest (2005) introduced a new hybrid solution, termed SOLAP technology, which fully integrates OLAP and GIS without requiring development efforts (commercially available as "JMap® Spatial OLAP Extension"). They make a clear distinction between a SOLAP application developed with any of the three approaches, and SOLAP technology which relies on the hybrid, fully integrated approach. In spite of such advances, today's SOLAP implementations remain basic and still face challenges to become more efficient.

The objective of the present chapter is to introduce the reader to the main advantages and challenges related to the use of spatial data and SOLAP. We begin with an overview of the particularities of spatial data to help readers to better assess the challenges that SOLAP development presents. After, we summarize the history of SOLAP applications and technologies as well as today's state-of-the-art. Then, we focus on the concepts, issues, challenges and solutions related to SOLAP. In the last sections, we discuss future trends and present concluding remarks. This content is presented from a geomatics engineering perspective; it is written for computer scientists who know the fundamental concepts related to spatial datacubes and OLAP.

BACKGROUND INFORMATION

With the recent evolution of geomatics sciences (geodesy-global positioning systems (GPS), photogrammetry, remote sensing, surveying, mapping and GIS), we gather today terabytes of land-related data everyday at a cost that is much lower that a decade ago. Mainstream applications using spatial data are appearing everywhere. Typically, they are very simple applications, they use spatial data obtained from a single source, they are out-of-date (one or more years old is the rule), they are incomplete and they show limited precision regarding the position of objects (in the order of tens of meters, and more). Although this is sufficient for most users (ex. tourists, news, routing), other users have more complex needs that require frequent updates, integration of data from different sources, integration of data from different epochs, integration of field measurements, integration of real time data, and so on. Insofar, most research in SOLAP has been done with the needs of the former group in mind. We can say that today's research community is succeeding into bringing spatial data into the OLAP arena. However, major challenges remain for the next several years in order to satisfy the needs of more advanced users. We still need to bring OLAP capabilities into the geomatics engineering arena.

Particularities of spatial data

Computer displays are flat, however the Earth is not. Furthermore, it is not a simple sphere nor a simple ellipsoid flattened at the poles. Earth's true shape looks more like a nice potato and it is scientifically defined as the geoid. The geoid is an equipotential surface that corresponds to the mean sea level. This physical model is the mathematical figure of the Earth as defined by its irregular gravity field. It is the model used by national mapping agencies to produce topographic maps upon which are based most thematic maps. It is more irregular than the ellipsoid of revolution because of the irregularities of the Earth surface (19,000 meters from the top of Mount Everest to the bottom of Mariana Trench) and because of the different densities associated with different types of minerals. The difference between the ellipsoid and the geoid can be up to 100 meters but we project our measurements on the ellipsoid to simplify the mathematics and to remain more stable over time (the geoid changes over time). Since the force of gravity is everywhere perpendicular to the geoid (not to the ellipsoid), our measurements (land-based or satellite-based) are influenced by the geoid. A slight vertical deviation of the measuring instrument may result into differences of hundred of meters when reporting a position from the geoid to the ellipsoid. The geomatics science that deals with the geoid and the ellipsoid is called physical geodesy and it provides the basis for all measurements.

Once we know the difference between the geoid and the ellipsoid, we must project the measured position to a flat surface such as a paper map or a computer display. This cannot be done without distortion, either of angles, areas or more typically of both at the same time. This has immediate impact on the shapes of objects, lengths, perimeters, areas and positions for example. To control these distortions, we use different map projections having different mathematical properties. Thus, from a unique position and shape on an ellipsoid, we obtain different shapes and positions on different maps made with different map projections. These differences may be up to hundreds of meters in some cases. The geoid-ellipsoid-map transfer process is illustrated by figure 1.

In spite of the existence of a standard international ellipsoid, many mapping agencies prefer to use a national or continental ellipsoid that better fits the surface of their country. They also use different map projections that minimize the distortions for the geometry of their country. Furthermore, different projects or organisations within a same country often use different map projections over a same zone depending of the total area to be covered by their maps. Selecting the most appropriate ellipsoid and map projection allows them to minimize (for the entire zone covered by a map series) the distortion between map measurements and the measurements made on the Earth (i.e. with regard to the geoid).

Furthermore, there are different spatial referencing systems to determine the position of objects on maps. One may use a latitude-longitude-height international ellipsoidal coordinate system, an x-y coordinate system based on a map projection, an x-y-z coordinate system from a 3D digital terrain model, a street address or street-intersection, a place name (dedicated or official such as a toponym), a distance-direction to a landmark, a route-direction-distance-offset linear referencing systems, and so on.

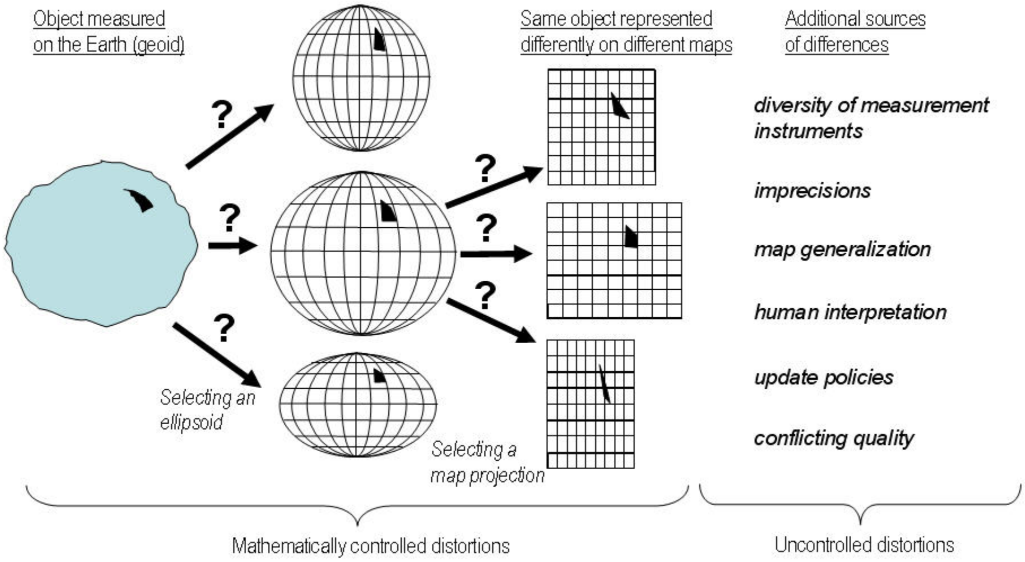

Figure 1: From one object measured on the Earth to different map representations.

The preceding differences can be controlled algorithmically and are currently handled by commercial software and interoperability standards. However, there remains other sources of distortion that cannot be controlled totally when dealing with multiple sources of data: diversity of measurement instruments, inherent imprecision of the measurement methods and tools used, data acquisition specifications that evolve over the years, limitations of the human interpretation of measured phenomena, independent data update policies, conflicting priorities over data quality, etc. The overall results are spatial data integration problems that cannot be avoided. Such problems happen for example when one integrates updates to an existing dataset (ex. original maps can have been made from aerial photographs but updates may be coming from field surveys required by municipal bylaws). They also happen when one integrates data from two adjacent maps made by two municipalities. They also take place when integrating different data collected independently for different purposes such as land use maps and utilities maps. They also occur when using real-time GPS vehicle tracking over a road map made from satellite imagery. Many more examples could be presented to explain why spatial data never fit together or with the reality! This creates major challenges for SOLAP as explained later in this paper.

In addition to the preceding issues, when one needs to have a more global cartographic view of a phenomenon, it is not possible to simply aggregate spatial data since the map or display becomes overcrowded and unreadable. One must rather use map generalization processes. According to Weibel and Dutton (1999), “Map generalization is responsible for reducing complexity in a map in a scale reduction process, emphasizing the essential while suppressing the unimportant, maintaining logical and unambiguous relations between map objects, and preserving aesthetic quality”. Every map, including the map made from source data, uses some level of generalization. By definition, a map is a model of only a subset of the reality where unnecessary details are eliminated and useful data emphasized while maintaining the map readable. Going from a large map scale to a smaller map scale worsens the situation. Categories of objects as well as individual objects are eliminated, others are replaced by a symbol of larger or smaller size, some are displaced, their shape is simplified, topological relationships may change, groups such as “building blocks” replace individual buildings where the density is too high, and so on. In other words, the content of every map may lie, the measurements made on every map may lie and a topological relationship on every map may lie. Taking this into account in a multi-resolution spatial datacube goes beyond the traditional topological concepts.

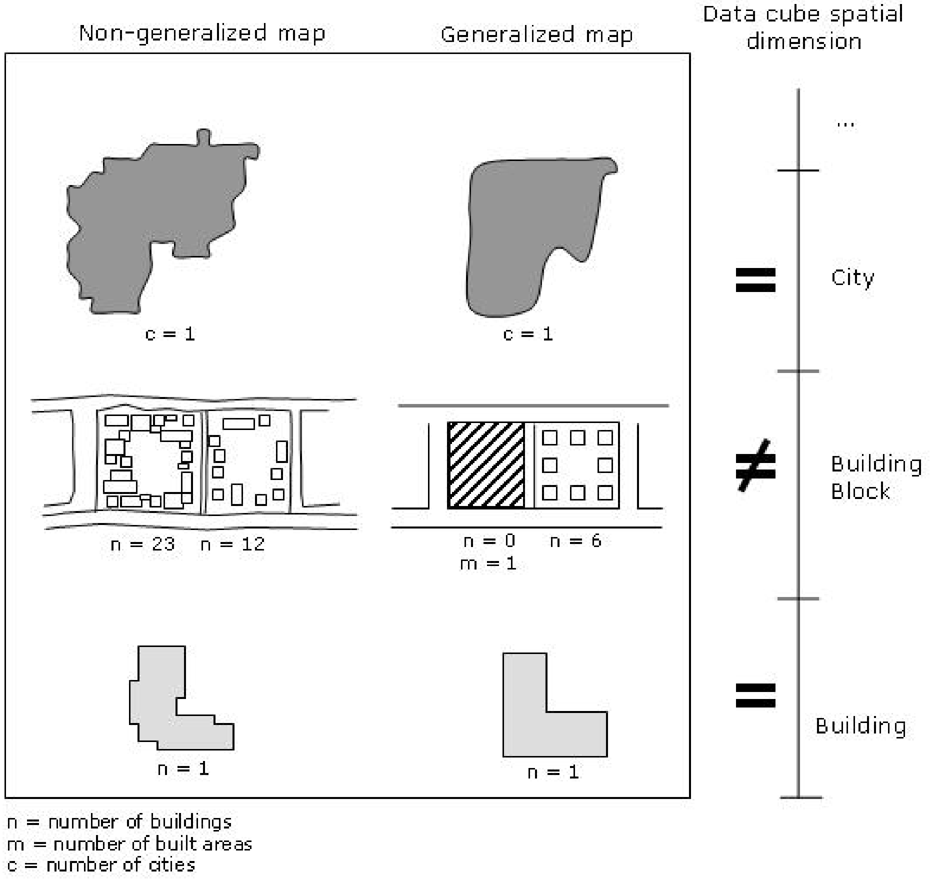

Figure 2: example of the spatial aggregation-generalization mismatch where aggregated data provide true data but unreadable map while generalized data produce readable map but inexact data.

Figure 2 shows an example of the impact of map generalization with regard to the spatial dimension of a datacube (a “spatial aggregation-generalization mismatch”). This introduces specific concerns for the interactive exploration of spatial data as seen later in this chapter.

From a geomatics point of view, spatial and temporal data are typically considered different from thematic data. Spatial and temporal data are reference data; they are used to locate phenomena in space and time rather than describing them. They are not intrinsic to a phenomenon like thematic data are. Since the human brain has built-in capabilities to use space and time, users intuitively rely on spatial and temporal data to integrate other data obtained from different sources. Such integration from different sources is typically performed without a priori planning since space and time are perceived by most to be universal reference systems. The preceding pages have shown that positioning objects in space is more complex than it appears, creating unexpected problems when one develops more advanced SOLAP applications. Time shows similar problems but to a much lower level of difficulty since it can be perceived as a one-dimensional universe with 0D (instant) and 1D (interval) events, as compared to the 3D spatial universe having 0D (point), 1D (line), 2D (surface) and 3D (volume) objects. From a computer science point of view, recent technical developments have reduced the need to maintain distinctions between spatial data and other data (Longley, Goodchild, Maguire & Rhind, 1999). Nevertheless, special care must always be taken when processing data from different maps and positioning technologies (ex. GPS) as it typically is the case when we build spatial data warehouses and spatial datacubes.

Evolution of SOLAP

In spite of a decade of research, testing and experimentations, it is only recently that SOLAP applications have been implemented into organisations for their daily decision-making. Examples in very diverse fields exist in Canada, France, United States and Portugal in particular. Furthermore, products supporting some SOLAP requirements have appeared on the market recently, either from key players such as SAS, ESRI, MapInfo, Business Objects and Cognos or from smaller innovative companies such as KHEOPS Technologies and ProClarity. These applications and technologies are still at their infancy but they already provide services that were not available previously to the users.

The term Spatial OLAP, or SOLAP, was coined by Bédard (1997) in parallel to the term Spatial Databases. Several research projects aiming at combining analytical databases and spatial databases were carried out since the mid-nineties. Pioneers from Simon Fraser University developed the GeoMiner prototype (Stefanovic, 1997) that included an efficient method for spatial datacube materialization (Han, Stefanovic & Koperski, 1998; Stefanovic, Han & Koperski, 2000). Other pioneers from Laval University (Bédard, 1997; Rivest, Bédard & Marchand, 2001) experimented varied combination of GIS and OLAP technologies with external users in different fields of application (Bédard et al., 2005) before developing the first commercial hybrid solution: JMap® Spatial OLAP Extension (Bédard, 2005). They developed several concepts, including new OLAP functions, spatio-temporal topological dimensions (Marchand, 2004), the use of raster representations of space for evolving datacubes (Miquel, Bédard & Brisebois, 2002) and integrating multiple-representations in spatial datacubes (Bédard & Bernier, 2002; Bernier & Bédard, 2005) for instance. Many research projects have built bridges between OLAP and GIS to facilitate the development of hybrid systems similar to the most recent commercial releases, such as GOAL (Kouba, Matousek & Miksovsky, 2000), SIGOLAP (Ferreira, Campos & Tanaka, 2001), SOVAT (Scotch & Parmato, 2005), GMLA Web Services (Silva, Times, Fidalgo & Barros, 2005) and CommonGIS (Hernandez, Voss & Gohring, 2005). A team from the University of Minnesota developed MapCube, a data structure and visualization tool for spatial datacubes (Shekhar, Lu, Tan, Chawla & Vatsavai, 2001). Another group from INSA-Lyon, in France, also developed a prototype of SOLAP application (Tchonikine, Miquel, Laurini, Ahmed, Bimonte & Baillot, 2005). In collaboration with the Laval University team, they worked on evolving dimensions (Body, Miquel, Bédard & Tchounikine,

2002) and highly heterogeneous data (Miquel et al., 2002). Fidalgo, Times, Silva and Souza (2004) have proposed a *GeoDWFrame* based on the star schema to facilitate the design of spatial dimensional schemas. Several Italian researchers have been active in fundamental research related to SOLAP. Pourabbas, Rafanelli, Ferri and others have published widely on PQL, a pictorial query language for spatial data using OLAP operators (Ferri, Pourabbas & Rafanelli, 2002; Pourabbas & Rafanelli, 2002; Pourabbas, 2003). Pourabas (2003) has presented the use of binding attributes to build a bridge while preserving the structure of both the spatial database and the OLAP datacube. Pestana is developing the concept of spatial dashboard based on SOLAP technology and collaborates with Laval team on conceptual modeling of spatial datacubes (Pestana, da Silva & Bédard, 2005). Other projects aim at improving spatial indexation, spatial aggregation or spatial operators (e.g. Gupta, Harinarayan, Rajaraman & Ullman, 1997; Han et al., 1998; Papadias, Kalnis, Zhang & Tao, 2001; Prasher & Zhou, 2004; Stefanovic et al., 2000; Wang, Pan, Ren, Cui, Ding & Perrizo, 2003; Zhang, Li, Rao, Yu, Chen & Liu, 2003; Zhou, Truffet & Han, 1999).

In spite of all this research activity, commercial solutions efficiently coupling OLAP and GIS appeared on the market only very recently. These solutions, some OLAP-centric, some GIS-centric, some hybrid, present only a subset of the desirable functionalities of a Spatial OLAP technology. Some are still limited to static map visualization of OLAP query results. Others require the storage of each potential individual map view on the server, thus affecting the update effectiveness of spatial data. These solutions present limitations with regards to interactive data manipulation and exploration through cartographic views. However, the main bottleneck still remains the building of spatial datacubes, especially when data come from different sources.

SOLAP CONCEPTS

Spatial OLAP (SOLAP) 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, 2004). Key to SOLAP concepts are multi-resolution spatial databases or data warehouses. Of particular interest is the 4-tier architecture for spatial data warehousing. The first tier represents the first data warehouse where integrated, homogeneous detailed data are stored. This first tier is very useful since the integration of spatial data from heterogeneous sources often requires human intervention. The second tier represents a second data warehouse where the results of the aggregation processes are stored. Since automatic map generalization is not fully automated and requires important human intervention, this second tier is necessary to store the results. The third tier is comprised of the datamarts, which can be further processed and organized according to a vertical view of the data (ex. within a range of map resolutions) or a horizontal view (ex. within a region or a department). The fourth tier includes the SOLAP clients that can add local information to the datamarts. Such architecture is particularly useful when the fusion of detailed source data represents important efforts that cannot be fully automated. Difficulties related to spatial data warehousing are found in (Bernier & Bédard, 2005).

The SOLAP concepts support the multidimensional paradigm and enriched data exploration based on an explicit spatial reference represented on maps. This explicit spatial reference can relate to dimensions and measures as SOLAP supports “spatial” dimensions and “spatial” measures. Three types of spatial dimensions can be defined: the non-geometric spatial dimensions, the geometric spatial dimensions and the mixed spatial dimensions (Bédard, Merrett & Han, 2001). In the first type of spatial dimension, the spatial reference uses nominal data only (ex. place names) as no geometry or cartographic representation is associated with the dimension members. It is the only type of spatial dimension supported by non-spatial OLAP. This type of spatial dimension is treated like other descriptive dimensions causing the spatio-temporal analysis to be potentially incomplete and the discovery of certain spatial relations or correlations between the phenomena under study to be missed by the analyst. The two other types of spatial dimensions aim at maximizing the potential to discover spatial relations and correlations that do not fit in predefined boundaries. The geometric spatial dimensions comprise, for all dimension members, at all levels of detail, geometric shapes (ex. polygons to represent city boundaries) that are spatially referenced to allow their dimension members (ex. New-York) to be visualized and queried on maps. The mixed spatial dimensions comprise geometric shapes for a subset of the levels of details. The members of the geometric and mixed spatial dimensions can be displayed on maps using visual variables that relates to the values of the different measures contained in the datacube being analyzed. Figure 3 presents examples of the three types of spatial dimensions.

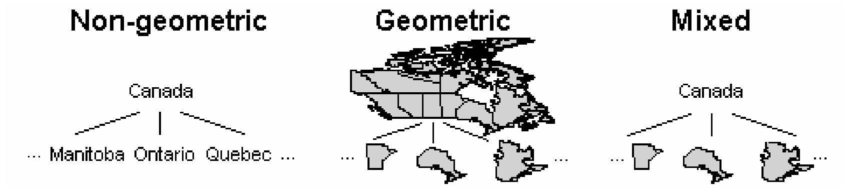

Figure 3: Three types of spatial dimensions: non-geometric, geometric and mixed spatial dimensions. After (Rivest, Bédard, Proulx & Nadeau, 2003).

Two types of spatial measures can be defined (Bédard et al., 2001; Han et al., 1998; Rivest et al., 2001; Stefanovik, 1997; Tchounikine et al., 2005). A first type is geometric. It is the set of all the geometries representing the spatial objects corresponding to a particular combination of dimension members from one to many geometric or mixed 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. A second type of spatial measure is numeric. It results from the computation of spatial metric or topological operators. Examples of this type of spatial measure could be “surface” and “distance” as well as “number of neighbours”. Figure 4 presents an example of the two types of spatial measures. A set of measures (spatial and non-spatial) organized according to a set of dimensions (spatial and non-spatial) form a spatial datacube.

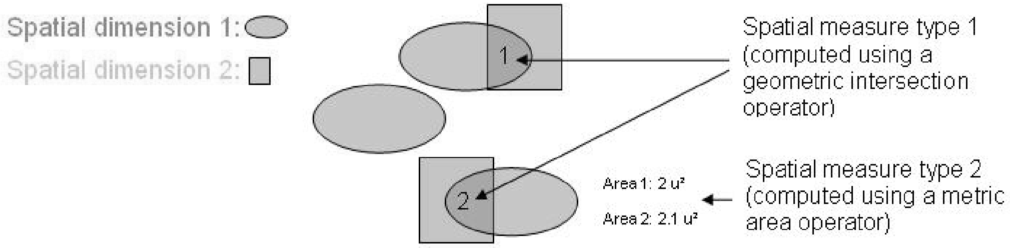

Figure 4: The two types of spatial measures supported in SOLAP tools.

In a SOLAP client interface, variants of the OLAP operators are used in order to take advantage of the spatial multidimensional data structure and of the different levels of detail of the spatial data. The general operators are drill-down, roll-up (or drill-up), drill-across, swap (or pivot) and slice and dice. These 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 (Rivest et al., 2003). *Thematic* operations allow the manipulation of thematic (or descriptive) dimensions, while keeping the same level of spatial and temporal granularities. *Temporal* operations allow the manipulation of temporal dimensions, while keeping the same level of thematic and spatial granularities. *Spatial* operations allow the manipulation of the spatial dimensions while keeping the same level of thematic and temporal granularities. Spatial operations can be executed directly by clicking on the elements (dimension members) shown on the maps.

SOLAP tools can support different types of views: tables, various types of diagrams and charts, and various types of maps. These include simple maps (i.e. single maps showing many geometric elements), multimaps (i.e. many maps, each related to a particular parameter, for example one map per year), 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 charts superimposed on the geometric elements of the map). The SOLAP views can support operator synchronization and graphical semiology synchronization. Figure 5 presents an example of a SOLAP display comprised of three views that use semiology synchronization (in a context of car accidents analysis).

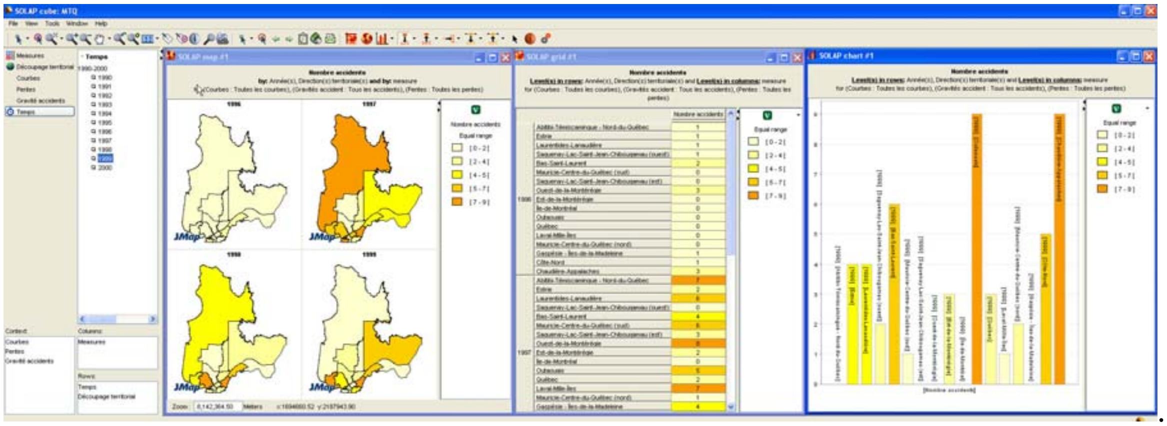

Figure 5: An example of use of semiology synchronization in a SOLAP.

The semiology synchronization helps to keep a visual homogeneity from one display to the other and from one navigation operation to the other. However, it involves taking care of potential collisions of graphical semiology rules since theoretically, the same rules do not always apply to all view types. Many more geographic visualization, or geovisualization, features can be incorporated. Geovisualization represents a research field in its own and can be defined as a private activity in which unknowns are revealed in a highly interactive environment. 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, 1994).

SOLAP tools may be used to implement a wide range of spatially referenced decision applications. For example, a road network management application may help to find in seconds without SQL queries the effects of variations in the annual average daily traffic on the average road conditions, or calculating the intervention costs (Rivest et al., 2001). In a similar way, it is possible to analyze the number and the gravity of car accidents according to their position on the road network, the characteristics of the road or the environment and the time period (Rivest, Gignac, Charron & Bédard, 2004). Another example is an environmental health application that allows investigating the relationships between health and environmental phenomena, like the incidence of respiratory diseases according to air quality measurements (Bédard, Gosselin, Rivest, Proulx, Nadeau, Lebel & Gagnon, 2003). Another example relates to the training of olympic-level speed skating athletes using GPS measurements, they use SOLAP to analyse their performances on various sections of a track according to various technical, mechanical and meteorological parameters (Veilleux, Lambert, Santerre & Bédard, 2004). In forestry, a 3D SOLAP application has also been recently implemented (Brisebois, 2004) as well as one for archeology (Rageul, 2004). These applications benefit from the three-dimensional aspect of space, that is the volume of the phenomena being studied. For instance, when archaeological excavation lots are represented as volumes, it is possible to navigate in the various stratigraphic units to compare the lots according to their color, granulometry, consistency, geographic and stratigraphic positions and the type of artefacts found (Fortin and Bédard, 2004).

SOLAP issues, challenges and recommendations

Although there remain computing challenges for SOLAP, our geomatics engineering perspective leads us to see the most crucial issues as the ones that relate to the management and processing of spatial referencing. We need to facilitate the flow of spatial data from the sources to the datacubes. When compared to traditional GIS research, research in spatial data warehousing, spatial OLAP and spatial datacubes requires to deal with more complex issues such as the efficient integration of time (which is ubiquitous in datacubes) with space (very few GIS databases are temporal), such as the production of spatial data at different levels of granularity for a same paper map or computer display size (on-the-fly automatic map generalization still remains a research topic), and such as the efficient integration of spatial data from heterogeneous and spatially divergent sources for instance (in spite of advances in interoperability, uncontrolled distortions cannot be resolved automatically). In particular, spatial aggregation and summarization often cannot be derived from detailed spatial data, requiring the use of smaller-scale maps from other sources and to automatically match corresponding spatial objects. Other challenges include: (1) better integration of spatio-temporal operators (topological and metric) to feed the datacube with spatial facts; (2) improved conceptual models of spatio-temporal datacubes; (3) development of a SOLAP design method to help facing the numerous spatio-temporal interrogations one may have and to better optimize spatial datacubes; (4) more advanced graphical semiology rules simultaneously compatible with maps, charts and tables; (5) enriched integrity constraints that combine spatial, temporal and aggregation constraints; (6) explicit ways to assess and display the estimated quality (internal and external) of information; (7) improve existing technologies for enriched SOLAP and better integration into spatial data production workflows.

FUTURE TRENDS AND CONCLUSION

Research related to spatial data warehousing and spatial OLAP has grown over a ten year period from the first ideas developed in a small number of isolated university laboratories to today's emergence of an R&D community. Researchers from several countries are addressing fundamental issues. Insofar, this community comes mainly from computer science departments. The geomatics community is only discovering the power of datacubes and OLAP. Rather, this community has looked into other directions such as geovisualization, advanced GIS and expert systems to better support spatial decision-making and geographic knowledge discovery. Looking at the issues of SOLAP from a geomatics engineering perspective is very promising. It brings a new level of challenges that relate to the very nature of spatial data and its use in multi-stakeholder environments. This enriches the concepts and technologies already available. In particular, it allows the integration of the early SOLAP solutions into the mainstream of spatial data production which is highly more complex than perceived at first sight. To further advance knowledge and to improve SOLAP applicability to complex interoperable environments, it is necessary to merge knowledge from the geomatics and the computer science communities. We expect that the most significant trends will emerge from this combination. From a scientific point of view, these trends would include the support of highly-efficient building of spatial datacubes (i.e. without human intervention), real-time SOLAP, mobile SOLAP, spatial dashboards and spatially-constrained data mining. From a commercial point of view, trends are likely to follow the typical evolution from bridging separate technologies (OLAP-centric or GIS-centric) into more integrated solutions (bidirectional bridges with common user interface) into fully integrated technologies that interoperate via web services and interoperate with spatial legacy systems.

AKNOWLEDGEMENTS

We recognize the financial support of Canada NSERC Industrial Research Chair in Geospatial Databases for Decision-Support and its partners (http://mdspatialdb.chair.scg.ulaval.ca/)

REFERENCES

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

Bédard, Y. (2004, April). *Amélioration des capacités décisionnelles des SIG par l'ajout d'un module SOLAP (Spatial On-Line Analytical Processing)*. Paper presented at the École Polytechnique Universitaire de Marseille, France.

Bédard, Y. (2005, May). *Integrating GIS and OLAP: a New Way to Unlock Geospatial Data for Decision-making*. Paper presented at the Location Technology and Business Intelligence Conference, Philadelphia, PA.

Bédard, Y., & Bernier E. (2002, July). *Supporting Multiple Representations with Spatial View Management and the Concept of "VUEL"*. Paper presented at the Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, ICA Comm. on Map Gen., Ottawa, Canada.

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., 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* (pp. 53-73). London: Taylor & Francis.

Bédard, Y., Proulx, M. J., & Rivest, S. (2005). Enrichissement du OLAP pour l'analyse géographique: exemples de réalisations et différentes possibilités technologiques. *Revue des nouvelles technologies de l'information*, B-1, 1-20.

Bernier, E., & Bédard Y. (2005, In press). Using a data warehousing architecture to combine automatic generalization and multiple representation for web-based on-demand mapping. In A. Ruas & W. McKinnis (Eds.), *Challenges in the Portrayal of Geographic Information: Issues of Generalisation and Multi Scale Representation*.

Body, M., Miquel, M., Bédard, Y., & Tchounikine, A. (2002). A Multidimensional and Multiversion Structure for OLAP Applications. In *Proceedings of the ACM Fifth International Workshop on Data Warehousing and OLAP* (pp. 1-6).

Brisebois, A. (2004). *Analyse du potentiel d'extension du concept SOLAP pour l'exploration des données spatiales 3D*. Unpublished Master's thesis, Laval University, Canada.

Buzan, T., & Buzan, B. (2003). *Mind Map, dessine-moi l'intelligence, 2^{è} édition*. Paris: Éditions de l'Organisation.

Ferri, F., Pourabbas, E., & Rafanelli, M. (2002). The syntactic and semantic correctness of pictorial configurations to query geographic databases by PQL. In *Proceedings of SAC 2002* (pp.432-437).

Ferreira, A. C., Campos, M. L., & Tanaka, A. (2001). An Architecture for Spatial and Dimensional Analysis Integration. In *Proceedings of SCI 2001, volume XIV computer Science and Engineering. Part II* (p. 392-395).

Fidalgo, R. N., Times, V. C., Silva, J., & Souza, F. F. (2004). GeoDWFrame: A Framework for Guiding the Design of Geographical Dimensional Schemas. In Pr*oceedings of DaWaK 2004, Lecture Notes in Computer Sciences 3181* (pp. 26-37).

Fortin, M., & Bédard, Y. (2004, October). *Développement d'un système de découverte des connaissances spatio-temporelles issues d'un chantier de fouilles archéologiques*. Paper presented at Géomatique 2004, Canadian Institute of Geomatics, Montreal, Canada.

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

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

Gupta, H., Harinarayan, V., Rajaraman, A., & Ullman, J. D. (1997). Index Selection for OLAP. In *13th International Conference on Data Engineering* (pp. 208-219).

Han, J., Stefanovic, N., & Koperski, K. (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).

Hernandez, V., Voss, A., & Gohring, W. (2005). Sustainable Decision Support by the Use of Multi-Level and Multi-Criteria Spatial Analysis on the Nicaragua Development Gateway , From

Pharaohs to Geoinformatics. In *Proceedings of the Fédération Internationale des Géomètres Working Week 2005 and GSDI-8* (pp.16-21).

KHEOPS. (2005). *JMAP Spatial OLAP*. Retrieved August 22nd from http://www.kheops-tech.com/en/jmap/solap.jsp.

Kouba, Z., Matousek, K., & Miksovsky, P. (2000). On Data Warehouse and GIS integration. In *Proceedings of DEXA 2000, Lecture Notes in Computer Sciences 1873* (pp. 604-613).

LGS Group Inc. (2000). *Analysis of Health Surveillance Business Intelligence Tools and Applications*, Final Draft.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. (Eds.). (1999). *Introduction. Geographical Information Systems: Principles and Technical Issues, 2nd Ed.* Wiley.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. (2001). *Geographic Information Systems and Science*. Wiley.

MacEachren, A. M. (1994). Visualization in modern cartography: Setting the agenda. In A. M. MacEachren & D. R. F. Taylor (Eds.), *Visualization in Modern Cartography*, 28(1), 1-12.

Marchand, P. (2004). *The Spatio-Temporal Topological Operator Dimension, a Hyperstructure for Multidimensional Spatio-Temporal Exploration and Analysis*. Unpublished doctoral dissertation, Laval University, Canada.

Miquel, M., Bédard, Y. & Brisebois, A. (2002). Conception d'entrepôts de données géospatiales à partir de sources hétérogènes, exemple d'application en foresterie. *Ingénierie des Systèmes d'information*, 7(3), 89-111.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Papadias, D., Kalnis, P., Zhang, J., & Tao, Y. (2001). *Efficient OLAP Operations in Spatial Data Warehouses* (Technical Report HKUST-CS01-01). Hong-Kong: Hong Kong University of Science and Technology, Department of Computer Science.

Pestana, G., da Silva M. M., & Bédard, Y. (2005, In press). Spatial OLAP Modeling: An Overview Based on Spatial Objects Changing over Time. In *Proceedings of the IEEE 3rd International Conference on Computational Cybernetics*.

Pourabbas, E. (2003). Cooperation with Geographic Databases. In M. Rafanelli (Ed.), *Multidimensional Databases: Problems and Solutions* (pp. 393-432). Idea Group Publishing.

Pourabbas, E., & Rafanelli, M. (2002). A Pictorial Query Language for Querying Geographic Databases using Positional and OLAP Operators. *SIGMOD Record*, 31(2), 22-27.

Prasher, S., & Zhou, X. (2004). Multiresolution amalgamation: dynamic spatial data cube generation. In *Proceedings of the fifteenth conference on Australasian database, ACM International Conference Proceeding Series, 52* (pp.103-111).

Rageul, N. (2004). *Développement d'une application d'exploration de données géospatiales comme support à la fouille archéologique*. Unpublished undergraduate report, INSA-Strasbourg, France.

Rivest, S., Bédard, Y., & Marchand, P. (2001). Towards better support for spatial decision-making : defining the characteristics of Spatial On-Line Analytical Processing. *Geomatica*, 55(4), 539-555.

Rivest, S., Bédard, Y., Proulx, M. J., & Nadeau, M. (2003, October). *SOLAP: a new type of user interface to support spatio-temporal multidimensional data exploration and analysis*. Paper presented at the ISPRS Joint Workshop of WG II/5, II/6, IV/1 and IV/2 on Spatial, Temporal and Multi-Dimensional Data Modelling and Analysis, Quebec, Canada.

Rivest, S., Gignac, P., Charron, J., & Bédard, Y. (2004, October). *Développement d'un système d'exploration spatio-temporelle interactive des données de la Banque d'information du ministère des Transports du Québec*. Paper presented at Géomatique 2004, Canadian Institute of Geomatics, Montreal, Canada.

Shekhar, S., Lu, C. T., Tan, X., Chawla, S., & Vatsavai, R. (2001). Map Cube: A Visualization Tool for Spatial Data Warehouses, In H. Miller & J. Han (Eds.), *Geographic Data Mining and Knowledge Discovery* (pp. 74-109). London: Taylor & Francis.

Scotch, M., & Parmanto, B. (2005). SOVAT: Spatial OLAP Visualization and Analysis Tool. In *Proceedings of the 38th Hawaii International Conference on System Sciences* (p. 142.2).

Silva, J., Times, V., Fidalgo, R., & Barros, R. (2005). Providing Geographic-Multidimensional Decision Support over the Web. In *APWeb 2005: 7th Asia-Pacific Web Conference, Lecture Notes in Computer Sciences 3399* (pp. 477-488).

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

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

Stefanovic, N., Han, J., & Koperski, K. (2000). Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. *IEEE Transactions on Knowledge Discovery and Data Engineering*, 12(6), 938-958.

Tchounikine, A., Miquel, M., Laurini, R., Ahmed, T., Bimonte, S., & Baillot, V. (2005). Panorama de travaux autour de l'intégration de données spatio-temporelles dans les hypercubes. *Revue des nouvelles technologies de l'information*, B-1, 21-33.

Veilleux, J. P., Lambert, M., Santerre, R., & Bédard, Y. (2004, October). *Utilisation du système de positionnement par satellites (GPS) et des outils d'exploration et d'analyse SOLAP pour l'évaluation et le suivi de sportifs de haut niveau*. Paper presented at Géomatique 2004, Canadian Institute of Geomatics, Montreal, Canada.

Wang, B., Pan, F., Ren, D., Cui, Y., Ding, Q., & Perrizo, W. (2003). Efficient OLAP operations for spatial data using peano trees. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery* (pp. 28-34).

Weibel, R., & G., Dutton. (1999). Generalising spatial data and dealing with multiple representations. In P. Longley, M. Goodchild, D. Maguire & D. Rhind (Eds.), *Introduction. Geographical Information Systems: Principles and Technical Issues, 2nd Edition* (pp. 125-155). Wiley.

Zhang, L., Li, Y., Rao, F., Yu, X., Chen, C., & Liu, D. (2003). An Approach to Enabling Spatial OLAP by Aggregating on Spatial Hierarchy. In *Proceedings of DaWaK 2003* (pp. 35-44).

Zhou, X., Truffet, D., & Han, J. (1999). Efficient Polygon Amalgamation Methods for Spatial (OLAP) and Spatial Data Mining. *Lecture Notes in Computer Sciences*, 1651, 167-187.