

Spatial OLAP: Open Issues and a Web Based Prototype

Sandro Bimonte, Anne Tchounikine, Maryvonne Miquel

LIRIS (Laboratoire d'InfoRmatique en Images et Systèmes d'information) UMR CNRS 5205

INSA, 7 Avenue Capelle, 69621 Villeurbanne Cedex, France

Name.Surname@insa-lyon.fr

INTRODUCTION

A Data Warehouse (DW) is a centralized repository of data acquired from external data sources and organized following a multidimensional model (Inmon, 1996) in order to be analyzed by On-Line Analytical Processing (OLAP) applications. OLAP tools provide the ability to interactively explore multidimensional data presenting detailed and aggregated data. The results of the analyses are the basis of strategic business decisions.

Spatial information is very often embedded in data but despite its significance, multidimensional models usually treat them as textual dimension. Integration of valuable spatial data in DWs leads to the definition of Spatial OLAP (SOLAP) (Rivest, 2005). SOLAP applications can address several and different domains: environmental studies, marketing, archaeology, epidemiology, etc...

Introduction of spatial data in multidimensional model raises major problems from implementation and theoretical point of view. This paper makes an inventory of some open issues in SOLAP. Then we propose a web based GIS-OLAP integrated solution supporting geographical dimensions and measures, and providing interactive and synchronized maps, pivot tables and diagrams displays in order to effectively support decision makers.

OLAP AND DATA WAREHOUSING

In multidimensional models, data is organized as an *n-dimensional cube* or *hypercube*. Figure 1 shows an example of a three-dimensional cube, which represents the sale of products in cities within a given time. Each dimension, i.e. cube's axes, can present a hierarchical structure which organizes data at different levels of details usually based on classification or generalization/specialization relationships. Facts are described by numerical values, called measures and stored in the cube's cells. (e.g. the number of items sold). Typical OLAP queries can involve multiple selections, nested group-bys and aggregated values, e.g. "Display the number of TVs sold per cities in USA and per month, and display the subtotals for each quarter". The result is a multidimensional and multi-levels table.

Data warehousing architectures are usually three-tier architecture (figure 2). The first tier is a warehouse server where data of interest is loaded after being extracted, cleaned, transformed and homogenized from operational legacy databases.

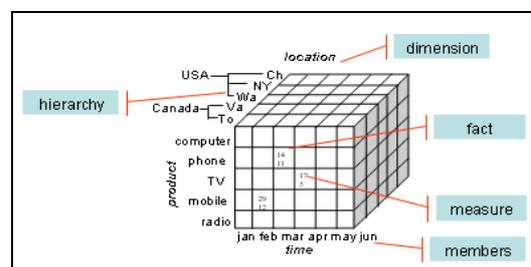


Figure 1: Sales Cube.

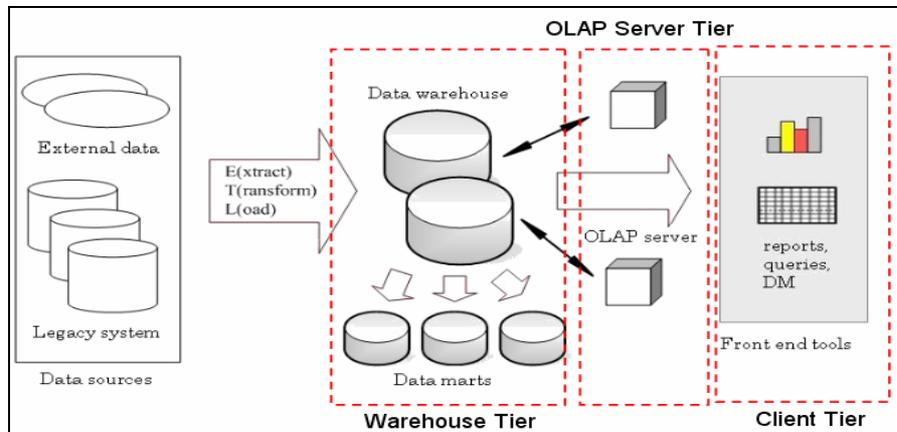


Figure 2: Data Warehouse Architecture.

In order to optimize data accesses, group-bys are pre-calculated by the 2nd tier (OLAP server) in the form of aggregates using one of the classical SQL aggregation operators applicable on the measures. OLAP operators then can be performed on the hypercube: the Roll-Up operator permits to climb the hierarchy meanwhile aggregating measures, Slice and Dice select a part of the cube and the Drill-Down operator reverses the Roll-Up.

The third level is an OLAP client which offers a user interface with reporting, interactive analysis and/or data mining tools. The most preferred data presentation paradigm is the Pivot Table, a 2D spreadsheet with associated subtotals and totals that supports viewing complex data by nesting several dimensions on the x- or y-axis and displaying data on multiple pages. As a prerequisite, OLAP operators have to be mapped into user interactions with the pivot table and/or diagram displays.

SPATIAL OLAP OPEN ISSUES

SOLAP implies a real rethinking of OLAP concepts. A SOLAP paradigm should define spatial measure with adapted aggregation functions, spatial dimensions and hierarchies, spatial OLAP algebra, suited GUI interface mixing tabular and cartographic features, etc...

Spatial Dimension

In multidimensional models, spatial data can be integrated into dimensions. An example of SOLAP application is the supervision of French departments' air pollution (figure 3).

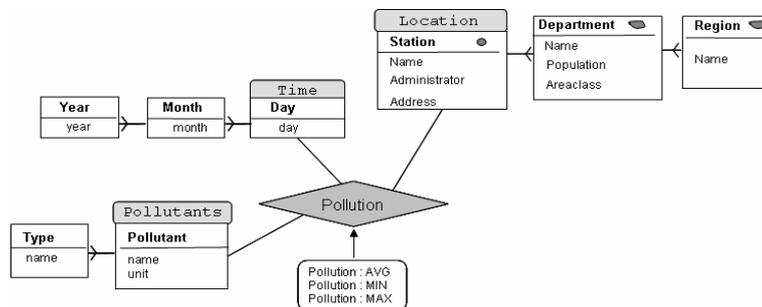


Figure 3: Spatial OLAP Pollution Application, with spatial geometric dimension.

Modeling spatial dimension

From a model point of view, one should be able to design a spatial dimension with textual attributes and/or spatial attributes (Rivest, 2005). The spatial attribute corresponds to a cartographic representation (i.e. point, line, and polygon). The multidimensional application of figure 3 addresses queries like: “*What is the average, min and max value for each pollutant for each department with an areaclass “cities and services” during 2000?*” As we can see, alphanumeric properties of spatial dimensions can be used to query the multidimensional cube.

One issue is to draw the hierarchy for the spatial dimension. Indeed, a spatial hierarchy is characterized by topological relations (i.e. inclusion, overlap) between members of the same and/or different spatial levels. This affects the accuracy of the aggregation process (Malinowsky, 2005). Moreover in order to introduce cartographic generalization in SOLAP, a support for managing complex hierarchies (i.e. non-covering) is unavoidable.

Visualization issues

Interactive maps and graphic displays are the main instruments to support a real and effective spatio-temporal analysis process. They reveal spatial trends, relationships and stimulate user’s thinking process (MacEachren, 2001). Indeed a cartographic representation of the multidimensional data permits to visualize measures on map and to discover geographical correlations between facts and members. Visualization is the most important feature and at the same time the main challenge of a SOLAP tool (Rivest, 2005). The user interface has to support more than one spatial dimension, maps mixing different spatial object representations (point, lines, regions, etc...), has to permit to navigate between cartographic and textual levels, and finally it has to support multidimensional analysis using maps, tabular and graphical data representation in an interactive and synchronized way. Synchronization means that an action (i.e. drill-down or roll-up) on one component (i.e. map) has to be “brought back” on the other ones (i.e. pivot table and graphic displays). The synchronization of the different visualization components is difficult as the expressive power of maps and pivot tables differs.

Moreover several semiotic problems have to be solved for the correct and relevant visualization of measures on spatial dimensions (bars, pies, etc...). How measures will be displayed on the map depend on several criteria: the nature of the measure (quantitative or qualitative?), the number of measures to be displayed, and the current representation of the spatial dimension (a point, a line or a polygon?). Thus ad-hoc semiotic rules for SOLAP are necessary, in order to grant a central and effective role in the decisional process to the map component.

Extended SOLAP algebra

It is obvious that for a full integration of spatial and multidimensional analysis, some new Spatial OLAP operators extending GIS operators have to be coined. For example, the clipping operator in a SOLAP tool can be used to define on-the-fly new spatial dimension members. Another way to use GIS operations is to use their result (for example a distance) as a predicate to slice or dice the cube.

Spatial and Geographical Measure

Spatial data can be the subject itself of the multidimensional analysis. The example figure 4 concerns the monitoring of the Venice Lagoon and uses the geometries of the “*unità barenali*” (Venice Lagoon sandbank zones) as spatial measures.

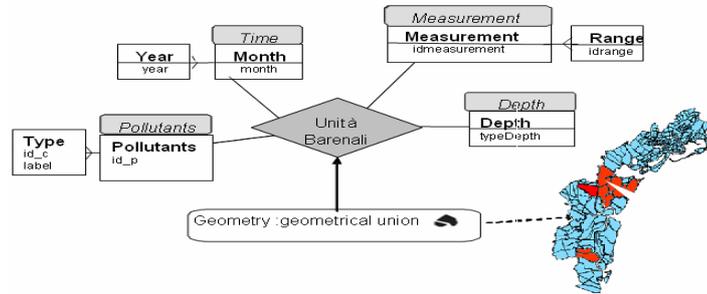


Figure 4: Spatial OLAP application of Venice Lagoon with Spatial Measure.

From spatial measure to geographical measure

In most of the state-of-the-art models (Malinowsky, 2005), (Marchand, 2003), (Rivest, 2005), (Stefanovic, 2001) the spatial measure is reduced to its geometric and/or metric part. In order to overcome this limitation, we introduce in (Bimonte, 2006a) the concept of geographical measure extending analysis to the alphanumeric attributes of a geographical object. Indeed, these attributes can be useful for the decision making process as they could hide precious information and/or characterize geographical locations. The geographical object “unità barenale” can better be described by 4 more descriptive attributes: name, index of salinity, list of plants and area (figure 5). This model allows responding to the following type of query: “return the entire zone, its plants and index of salinity where “metals” measured at depth “high” has a value of [0.05-0.07ng/l]?”

Modeling geographical measures implies a reformulation of classical formal multidimensional models and especially to re-define the aggregation process that must be applied on each attribute of the geographical measure. The attributes of a geographical object are inter-dependant and so are aggregation functions. For this reason a deeper investigation of additivity issues in SOLAP is needed.

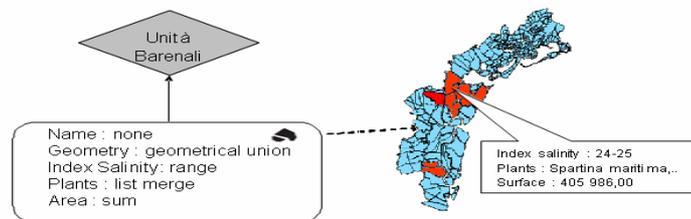


Figure 5: Spatial OLAP application of Venice Lagoon with Geographical Measure.

Visualization issues

From the visualization point of view, new paradigms are necessary to visualize geographical measures, combining pivot tables and advanced geovisualization techniques, by replacing the textual data of the cells with maps. This leads to several problems: the size of the pivot table and the maps, the correct visualization of a set of geographical measures in the same cell, etc... Thus in order to preserve the same expressive power of the pivot tables for numerical and geographical data, pivot table cells should include particular visualizations techniques for the spatial and alphanumeric information (i.e. space-time cubes (Bimonte, 2006b) and/or choremes representations).

Spatial OLAP Tools

Several Spatial OLAP tools have been developed until now. In (Rivest, 2005) the authors describe a SOLAP client tool which supports a tabular data representation, 7 different types of synchronized diagrams and maps composed by visual variables and maps superimposed with graphical diagrams.

Some efforts like (Scotch, 2005) try to combine data mining and SOLAP functionalities. (Silva, 2006) provides a web based SOLAP tool whose main feature is the usage of geographical web-services for multidimensional databases. Finally in (Hernandez, 2005) authors following an Exploratory Spatial Data Analysis approach, present CommonGIS: A powerful tool for interactive visual geo-analytics, extended to process hierarchical multidimensional data from OLAP warehouses.

Jmap (Jmap) and OLAP for ArcGIS (ESRI) are two commercial solutions which permit to integrate OLAP and GIS functionalities in a unique framework. JMap is a web-based solution.

This proliferation of SOLAP solutions confirms the importance and the value of this decision support technology, but at the same time the lack of a general framework establishing the mandatory functionalities, the spatial multidimensional query language and the “most-well suited” visualization techniques of a SOLAP tool.

GeWolap FOUNDATIONS

In this section we describe our web-based SOLAP prototype GeWolap. GeWolap is an extensible GIS-OLAP integrated solution supporting geographical dimension and measures, and based on a formal multidimensional model named GeoCube (Bimonte, 2006a). In this paper, we focus on the management of geographical dimensions. Geographical measures are detailed in (Bimonte, 2006c). GeWolap support two of the mandatory visualization issues as described in the previous section: Mixing different spatial object representations and components’ synchronization. Even if in our scenario of use, the members of spatial dimensions are always topological contained in the members of the less detailed spatial level, GeWolap allows overlapping spatial members.

GeWolap is based on a three tier architecture consisting of an Object-Relational DBMS (ORACLE) supporting spatial data, an OLAP Server (Mondrian), and a front-end client coupling an OLAP and a GIS web client. Mondrian (Mondrian) provides an easily customizable, full featured OLAP server (aggregate tables, user defined functions and complex hierarchies, MDX support). The client tier is implemented using JPivot (JPivot) to provide JSP pages and MapXtreme Java (MapXtreme) to support map visualization and interaction.

The User Interface

The client tier is aimed at providing an interactive environment which graphically encapsulates the structure of the multidimensional spatial application and translates interactions with the visual interface into operators. The user can interact by means of synchronized map and pivot table in order to navigate in the spatial data warehouse and access to several GIS and OLAP functionalities using only few mouse clicks. Moreover it is important to underline that as GeWolap supports numerical and also geographical measures, the user interface adapts its functionalities according to the particular type of application.

The GUI is composed of 2 main panels (figure 6):

- the *Control Panel* contains 3 tabs windows (figure 7):

- *Navigation* lists all available cubes classified per project
- *Display/Export* permits to select the visible pivot table, graphic display (figure 7) and/or map and allows export in various formats.
- *Analysis* provides a Cube Browser displaying dimensions, levels and members. It is used in order to map the hierarchies on table axes and to express selection of members to slice or dice the cube.
- *Data Panel* is composed by 3 frames and associated tool bars for each type of display (*pivot table* with associated *OLAP toolbar*, *diagram display*, and map with associated *SOLAP-GIS toolbar*).

The *OLAP Toolbar* (figure 8) provides functionalities in order to modify the visualization of the pivot table and triggers the desired OLAP operators: *drill-down replace*, *drill-down position*, *expand-all*, *drill-through*. The *drill-down replace* enables drilling from one pointed member to its child members in the dimension hierarchy, hiding the parents, whereas the *drill-down position* shows the parents. The *expand-all* operator enables drilling from all visible members in the table to child members.

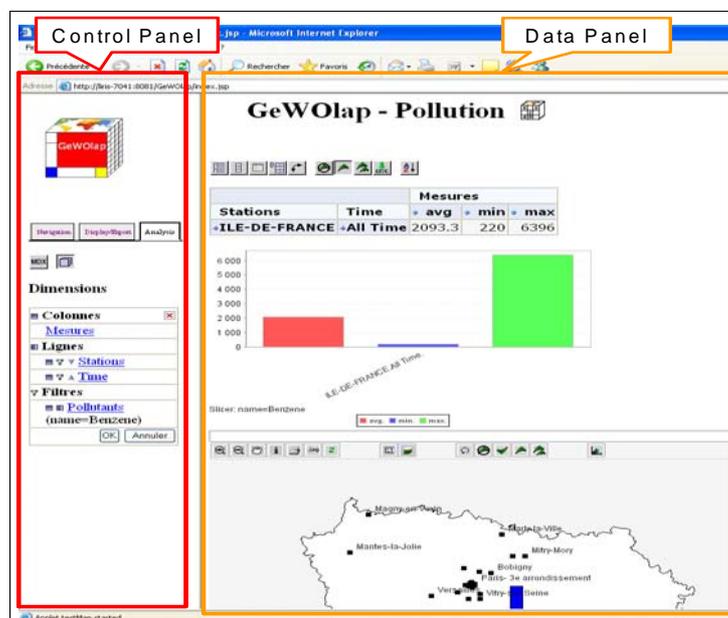


Figure 6: GeWolap Control Panel and Data Panel.



Figure 7: Control Panel: Navigation, Display/Export and Analysis windows tabs.

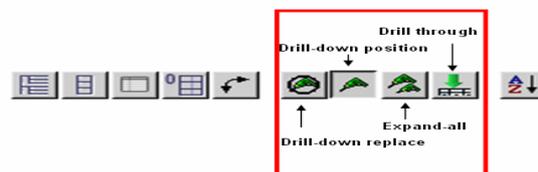


Figure 8: Data Panel: OLAP Toolbar.

The *Spatial OLAP-GIS Toolbar* (figure 9) provides:

- GIS Functionalities: *Zoom in/out*, *Pan*, *Retrieve* of alphanumeric information of geographical objects, *Map print*, *Map export*, *Map refresh*, *Rule tool* (which calculates distance between points) and *Control Layer* which permits to add layers.
- Spatial OLAP operators: *Spatial drill-down replace*, *Spatial drill-down position*, *Spatial roll-up*, *Spatial expand-all* are the spatial flavored OLAP operators. *Measure displays* permits to select the type of diagram used to represent the measures on the map.

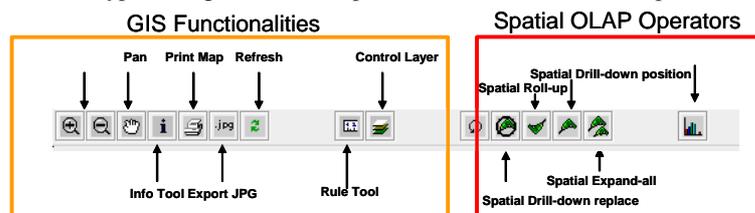


Figure 9: Spatial OLAP Toolbar.

NAVIGATING THROUGH A CUBE WITH GEWOLAP

In this section we illustrate the navigation process with GeWolap using the cube of figure 3.

OLAP Operators and Functionalities

The map and the pivot table represent the same multidimensional information. In order to visualize measures at the Department level, the user can choose the OLAP *drill-down position* operator, and click on the member ILE-DE-FRANCE in the pivot table. This drill-down operation triggers the generation of a map that shows the three measures for each department of the region

(figure 10). Synchronization of map and pivot table requires particular attention: Even if in the pivot table the measures associated to the region Ile-de-France are always present, they cannot be shown in the map as a region covers its departments.

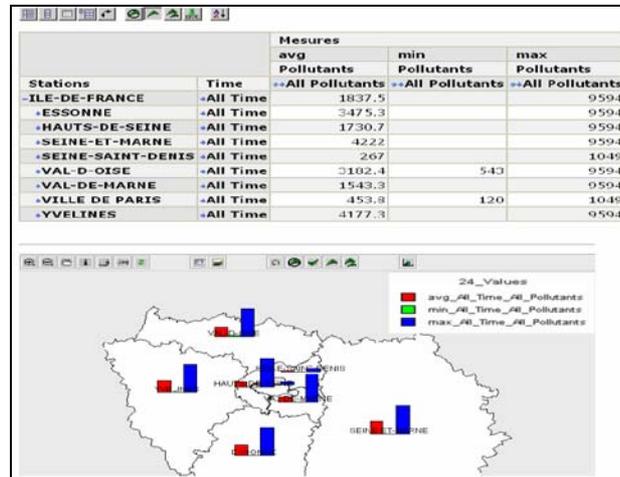


Figure 10: Drill-down position: avg, min and max measures grouped by Departments of Ile-de-France.

The user can select a particular measure using the Cube Browser in the Analysis window. The result is the new map-tabular configuration shown in figure 11 where the measure (here average pollution) is displayed using a theme range.

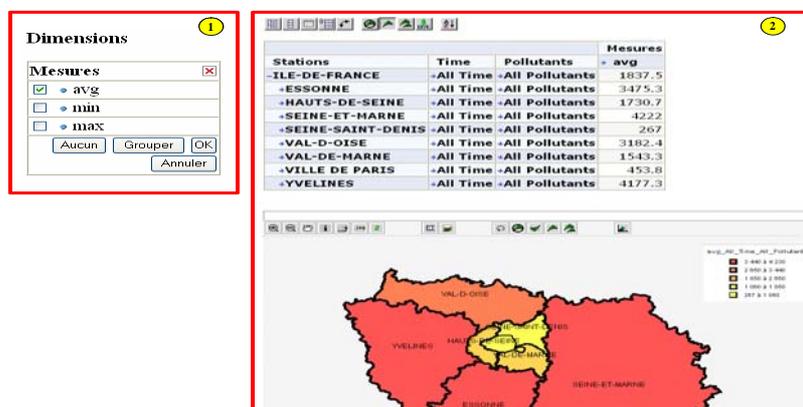


Figure 11: Selection of the avg measure using the Cube Editor and visualization of the avg measure grouped by Departments of Ile-de-France.

If interested in measure values for each class of pollutant for the particular department ESSONE, the user drills on the All Pollutants member on the pivot table. This action does not change the level of the spatial dimension but changes the measures shown on the map (figure 12).

Indeed the map legend mentions 3 measures:
 avg_All_time_all_pollutants, avg_All_time_primary_pollutants and

avg_All_time_secondary_pollutants, even if the pivot table shows always one measure. The pivot table and the map show the same information and they are always synchronized.

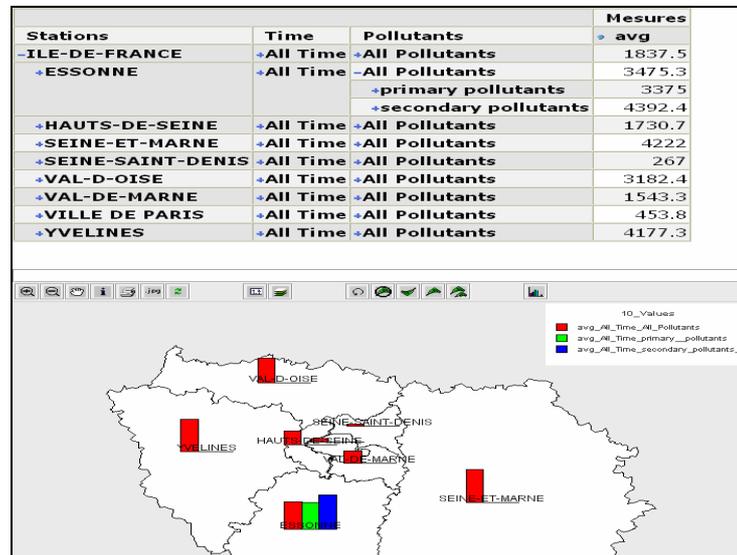


Figure 12: Drill-down position: avg measure grouped by Class of Pollutant and Departments of Ile-de-France.

Spatial OLAP Operators and GIS Functionalities

Let's show now the interaction from the map to the pivot table. Suppose that the user wishes to "see" the measures for the stations of VILLE DE PARIS. By pointing the mouse on that department s/he can apply the spatial drill-down position operator and then a Zoom In (figure 13).

In GeWolap GIS functionalities are provided too, as for example the choice of different visual representation for measures on map and for the departments (figure 14) or the addition of layers: Figure 6 represents the measures for Benzene Pollutant with a layer representing the stations.

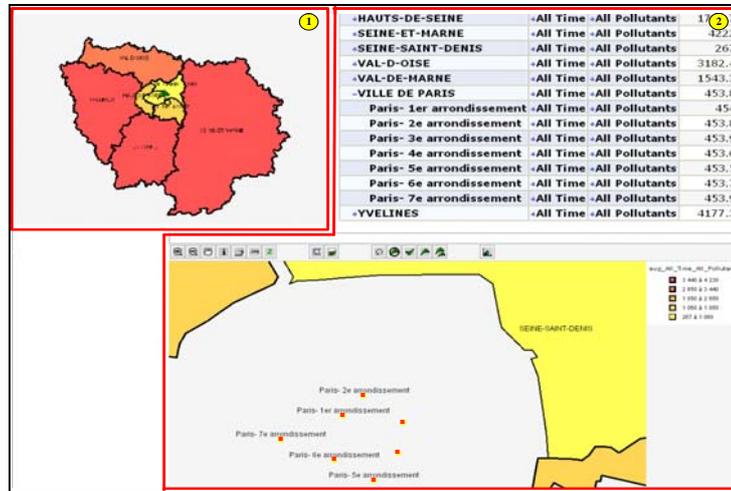


Figure 13: Spatial drill-down position and Zoom In: avg measure for each Station of VILLE DE PARIS.

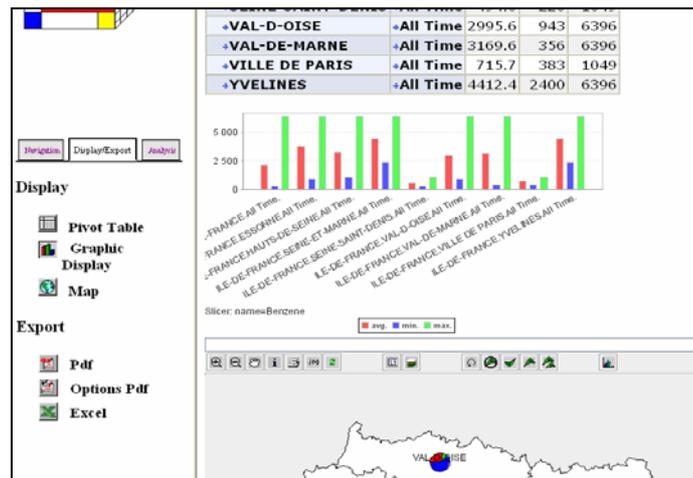


Figure 14: Diagram Display.

CONCLUSIONS AND FUTURE WORK

This work presents a panorama of SOLAP open research issues and a web based SOLAP prototype which tries to combine in a single framework the fundamental SOLAP features.

GeWolap supports spatial dimensions with numerical measures, and geographical measures. The user interface permits to explore the multidimensional spatial data by means of interactions with the synchronized map and pivot table components and combines GIS and OLAP functionalities.

Our current work addresses the introduction of other GIS functionalities and new Spatial OLAP operators extending classical GIS operators (for example buffer or clipping). We also investigate the definition of some SOLAP query patterns taking into account the number, the type and the current

representation of dimensions, measures types and aggregation functions involved in the query to help the choice of the most appropriate cartographic representation of SOLAP queries results.

BIBLIOGRAPHY

- Bimonte, S., Tchounikine, A., Miquel, M.: GeoCube, a Multidimensional Model and Navigation Operators Handling Complex Measures: a Spatial OLAP Application. In Proceedings of the 4th Biennial International Conference Advances in Information Systems, Izmir, Turkey, October 2006. LCNS 4243, Springer-Verlag Berlin Heidelberg: 100-109, 2006.
- Bimonte, S., Di Martino, S., Ferrucci, F., Tchounikine, A.: GeOlaPivot Table: a Visualization Paradigm for SOLAP Solutions. In Proceedings of the VLC Workshop of 12th International Conference on Distributed Multimedia Systems, Grand Canyon, USA: 181-186, 2006.
- Bimonte S., Wehrle, P., Tchounikine, A., Miquel, M. GeWOlap: a Web Based Spatial OLAP Proposal, Second International Workshop on Semantic-based Geographical Information Systems (SeBGIS'06), Montpellier, France, 29-30 October, Springer-Verlag Berlin Heidelberg: 2006
- ESRI, <http://www.esri.com/software/arcgis/extensions/olap/about/overview.html> as of 15/1/2007.
- Hernandez, V. Voss A. Göhring, W. Hopmann, C.: Sustainable decision support by the use of multi-level and multi-criteria spatial analysis on the Nicaragua Development Gateway. In Proc. 8th International Conference on the Global Spatial Data Infrastructure, Egypt: 2005.
- Inmon, W. H., 1996 The data warehouse and data mining. Commun. ACM, 39, 11, pp. 49-50.
- JMap, <http://www.kheops-tech.com/en/jmap/solap.jsp> as of 15/1/2007.
- Jpivot, <http://jpivot.sourceforge.net/> as of 15/1/2007.
- MacEachren A. M., 2001 An evolving cognitive-semiotic approach to geographic visualization and knowledge construction. Information Design Journal, 10,1, pp. 26-36.
- Malinowski, E., Zimányi, E.: Spatial Hierarchies and Topological Relationships in the Spatial MultiDimER model. In Proc. of the 22nd British Nat. Conf. on Databases, BNCOD22, LNCS 3567, Springer: 7-28, 2005.
- MapXtreme Java, <http://extranet.mapinfo.com/products/Overview.cfm?productid=1162> as of 15/1/2006.
- Marchand, P., Brisebois, A., Bédard, Y., Edwards G., 2003 Implementation and evaluation of a hypercubebased method for spatio-temporal exploration and analysis. Journal of the International Society of Photogrammetry and Remote Sensing, 59, pp. 6-20.
- Mondrian, <http://mondrian.pentaho.org> as of 15/1/2006.
- Rivest, S., Bédard, Y., Proulx, M.-J., Nadeau, M., Hubert F., Pastor, J., 2005 SOLAP: Merging Business Intelligence with Geospatial Technology for Interactive Spatio-Temporal Exploration and Analysis of Data. Journal of International Society for Photogrammetry and Remote Sensing (ISPRS) 60, 1, pp. 17-33.
- Scotch, M., Parmanto, B.: SOVAT: Spatial OLAP Visualization and Analysis Tool. In Proceedings of HICSS-38, Waikoloa, IEEE: 2005.
- Silva, J., Times, V., Salgado, A.: An open source and web based framework for geographic and multidimensional processing. In Proceedings of Symposium on Applied Computing, ACM press: 63-67, 2006.
- Stefanovic, N., Han, J., Koperski K., 2000 Object-Based Selective Materialization for Efficient Implementation of Spatial Data Cubes. IEEE TKDE, 12, 6, pp. 938-958.