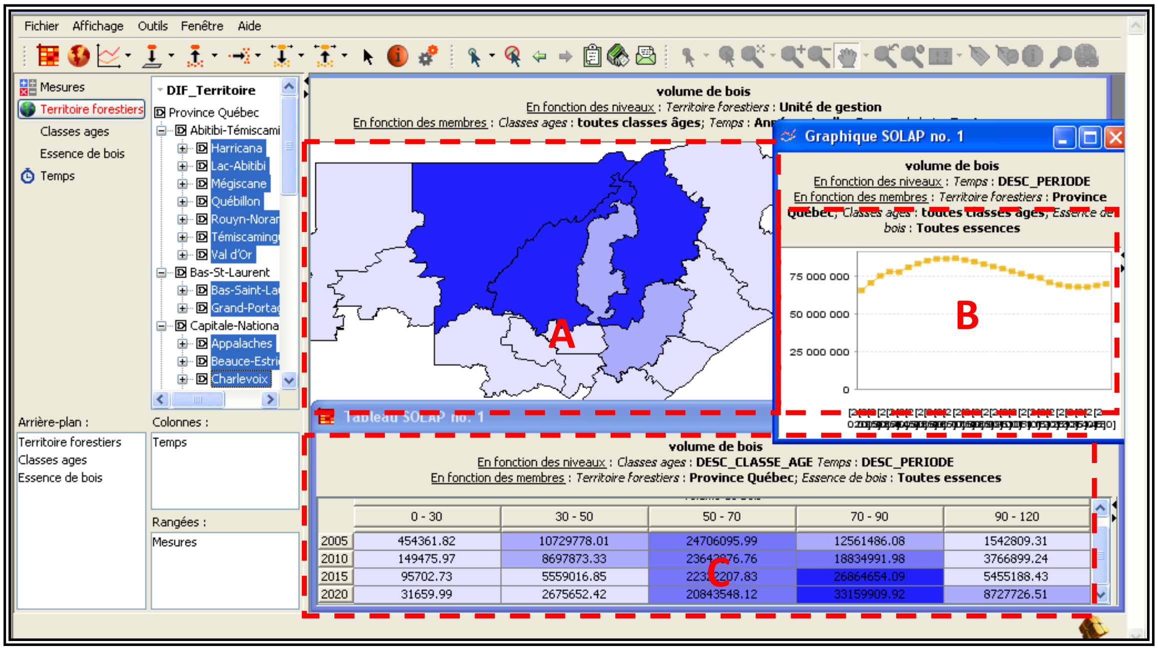
Interactive Exploration of Multi-granularity Spatial and Temporal Datacubes: Providing Computer-Assisted Geovisualization Support

Véronique Beaulieu[1] & Yvan Bédard[2]
Laval University
Centre for Research in Geomatics (CRG)
Canada NSERC Industrial Research Chair in Geospatial Databases for Decision Support
[1] veronique.beaulieu.3@ulaval.ca
[2] yvan.bedard@scg.ulaval.ca
http://sirs.scg.ulaval.ca/yvanbedard

This paper addresses issues of geovisualization regarding the process of interactive multi-granularity spatial and temporal knowledge discovery for decision-making. In such an interactive context, analysts expect to easily and instantly get knowledge out of spatial data the same way they can do it nowadays in the field of Business Intelligence for non-spatial data (e.g. using OLAP: On-Line Analytical Processing). Such knowledge must pop out from pertinent combinations of maps, tables and charts and immediately attract the analyst's attention. The power of envisioning data is necessary in the first phase of knowledge construction to stimulate the abduction and induction processes (Gahegan et al., 2001, Andrienko and Andrienko, 2000, Dyke and al, 2005), i.e. when only vague *a priori* knowledge exists and the analyst anticipates the construction of potential hypothesis. Such need requires specific technologies and visualization strategies that allow the analyst to go as fast as his train of thought requires. In other words, the analyst must focus on "seeing something", i.e. on "what" he's looking for (e.g. trends, comparisons, contextualisation) rather than focusing on "how to obtain it". The analyst should not (1) need to master a query language or to be an expert in cartographic visualization, (2) wait more than Newell's cognitive band of 10 seconds (Newell, 1990) before getting a complete response, whatever the level of aggregation involved or the complexity of cross-tabulations, (3) need to "read" the details of maps, tables or charts and extract himself the global picture, (4) need to synchronize the behaviour of maps, tables and charts during an interactive exploration, (5) struggle with incompatible visual variables amongst maps, tables and statistical charts, (6) and so on.

The technology required to support this level of spatial and temporal multi-granularity interactivity exists today. For example, it is possible to combine GIS or map visualisation technology with existing business intelligence (BI) technologies called OLAP (N.B. such technologies are also called *analytical technologies* in the BI literature). The result of such a combination has been coined SOLAP (Spatial OLAP) in the literature (Bedard et al, 2007). Several authors have defined the SOLAP underlying concepts (e.g. Malinowski and Zimanyi, 2008; Proulx et al, 2008; Bedard and Han, 2008; Scotch et al, 2007; Rivest et al, 2005) and it is nowadays being offered commercially by several solution providers such as Oracle, SAS and JMap (see http://www.spatialbi.com for more examples). Its main strength for the analysts is the high level of interactivity offered to explore the data resulting from various aggregation methods for different levels of abstraction and different themes (e.g. region: city->county->state->country; period: day->month->year; product: coat->winter clothes->clothing->all) and to cross-tabulate such themes at whatever level of data granularity (e.g. Canada-2006-coat = 124,000 sales; Montreal-2006-clothing = 2,423,000 sales).



volume de bois
En fonction des niveaux : *Territoire forestiers* : **Unité de gestion**
En fonction des membres : *Classes ages* : **toutes classes âges**; *Temps* : **An**

volume de bois
En fonction des niveaux : *Temps* : **DESC_PERIODE**
En fonction des membres : *Territoire forestiers* : **Province Québec**; *Classes ages* : **toutes classes âges**; *Essence de bois* : **Toutes essences**

volume de bois
En fonction des niveaux : *Classes ages* : **DESC_CLASSE_AGE** *Temps* : **DESC_PERIODE**
En fonction des membres : *Territoire forestiers* : **Province Québec**; *Essence de bois* : **Toutes essences**

	0 - 30	30 - 50	50 - 70	70 - 90	90 - 120
2005	454361.82	10729778.01	24706095.99	12561486.08	1542809.31
2010	149475.97	8697873.33	2364[illegible]76.76	18834991.98	3766899.24
2015	95702.73	5559016.85	223[illegible]2207.83	26864654.09	5455188.43
2020	31659.99	2675652.42	20843548.12	33159909.92	8727726.51

Figure 1 : Example of an actual SOLAP interface showing the spatial distribution of the volume of wood harvested with a map (A), the temporal evolution with a line graph (B), and a table of cross-tabulated data (C).

Other interesting technologies have also been developed to explore spatial and spatio-temporal data visually. Among them, we find CommonGIS (Andrienko and Andrienko, 2003; Slocum and al, 2005) and GeoVista (http://www.geovista.psu.edu). The main difference between such technologies and SOLAP is that SOLAP emerged from the Business Intelligence paradigm and consequently relies on the use of an analysis-oriented database structure (called spatial datacube) and their operators (e.g. drill-down, roll-up, drill-across, swap, slice).

In spite of such theoretical and technological advances, the proper use of geovisualization concepts to better support the interactive nature of analysis offered by SOLAP technology has not been explored insofar. Consequently, in spite of having the tools supporting interactive geographic knowledge discovery, there remains the need to introduce the concepts underlying the proper display of information that would accelerate the discovery of certain spatial and temporal trends, clusters, thresholds, correlations, hierarchical relationships, and so on. Hence, this research aims to integrate geovisualization concepts with SOLAP concepts (Spatial On-Line Analytical Processing).

The first step toward such integration is to better understand the involved processes: knowledge discovery, thematic mapping, and statistical charting. How do humans perceive relationships, clusters, differences? What are the key variables involved in the spatial, temporal and multi-granularity visualizations? How do cartographers and statisticians communicate to the readers the intended information? etc. Whether we are in an interactive analytical context or not, the proper use of visual variables is a key element as demonstrated in an abundant literature (Bertin, 1973; Béguin and Pumain, 2003, MacEachren, 1994). In other words, graphic semiology or rules are important to produce maps, tables and statistical charts that are expressive and convey the desired message. Moreover, knowledge discovery is well supported with interactive multiples views linked to each other. Multiple ways to look at data brings different perspectives, each representation highlighting different aspects of data. No “ideal” visualization would fit all the analyses, but combination of different usages allows envisioning different aspects of data. Hence, seeing data with multiple views helps constructing a general idea of trends and patterns that exist within them. (MacEachren, 1995; Rousseau et Fortin, 1992; Ware, 2000). Geovisualization based on (carto)graphic and knowledge discovery processes help highlighting facts as much as they help avoiding misinterpretations of data. This creates new challenges that have been addressed by the geovisualization community and which also need to be addressed by the SOLAP community to better support the geospatial visual analyses for discipline experts.

In this paper, we propose a solution based on the selection of key dimensions involved in building expressive SOLAP-based maps, tables and charts. In other words, we propose a solution to better support the interactive multi-granularity geographic knowledge discovery process typical of SOLAP applications. Such combinations of dimensions happen to be fit naturally with the nature of datacube structures as used in BI analytical technologies. They result in the possibility to suggest a list of potential parameters that guide the selection of more efficient (carto)graphic representations of data.

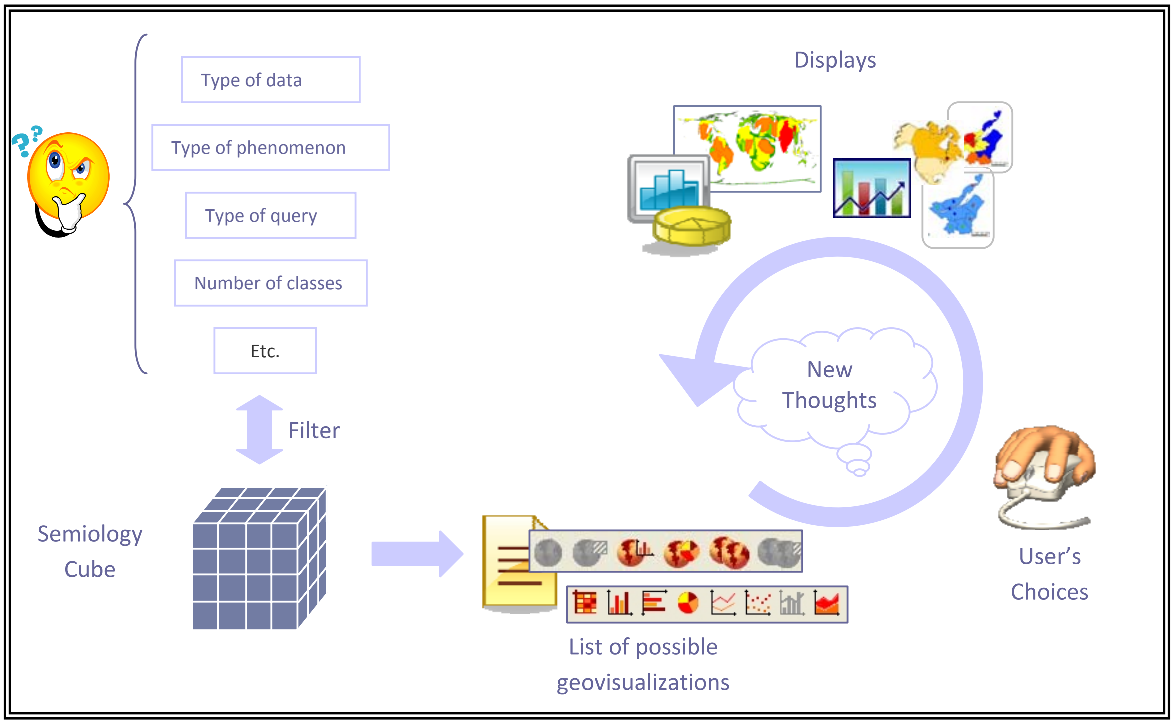


Figure 2: Role of the semiology cube in the user's process of exploration of data

These dimensions and the parameters determined by the combination of dimensions are predefined by specific rules and the results are stored in the datacube. Among the dimensions making the semiology datacube, we identified the type of data (nominal, ordinal, numeric), the type of query (or tasks according to Andrienko and Andrienko, 2005), phenomenon type (abrupt/smooth, discrete/ continuous), number of classes, etc. Among the parameters, we suggest templates for different types of statistical chart or map (e.g. Bertin's histogram or multiple link graph; cartogram or choropleth map) and visual variables (size or value, hue or form, etc.). The datacube provides a rich knowledge base for suggesting appropriate geovisualization solutions directly from the data being analysed and user's manipulations. When coupled with SOLAP technology, the semiology datacube enables SOLAP to offer correct, expressive and tailored geovisualization possibilities (an improvement over today's solution which allows all technologically supported geovisualizations, whether they make sense or not, thus requiring users or database administrators good knowledge of semiology and time-consuming configuration of representations). As a result, the analyst doesn't have to wonder how to visually present the data he wants to interpret, thus better keeping his train of thought. He simply has to choose the data he wants to see and choose one or several correct representations that the semiology datacube offers him. Furthermore, selecting amongst the proposed different combinations (or proposed templates) of visual variables may prove useful to help discover trends, clusters, evolutions, etc. as one combination may be more expressive than another for the observed data. This solution also removes the need to restrict the representations of data to solely those selected by the database administrator.

The proposed solution offers multiple advantages. First of all, it is amenable to a computer system as a stand-alone application as much as a part of a SOLAP implementation. Second, the parameter definition stored in the semiology datacube may respect the OGC standards for interoperability concerns. Third, since the visual parameters proposed by the datacube become readable by other technologies, the integration of the supported geovisualization concepts can be exported to other geospatial technology than SOLAP, for example to GISs and map viewers. Fourth, because multiples views proposed by the data cube could be readily built and accessed, the analyst can easily take a look at data in way that would never have come to his mind or that would have required too much time to build. The visual analysis would then bring different perspectives and different thoughts, enhancing further the exploration of data. Fifth, maximum flexibility is given to the user who may choose between the proposed geovisualization parameters during the exploration steps. Finally, flexibility is also given to designers or cartographers in a specific context to add new design into the cube. Once the new designs with its parameters are stored in the datacube, there is no need to redesign it for different data.

As far as we know, this research covers for the first time the integration of geovisualization concepts into SOLAP technology. Furthermore, it appears that the datacube structure underlying SOLAP technology offers an innovative way to take into consideration different dimensions of semiology rules and to use OLAP operators to extract the most appropriate geovisualization parameters for the selected dimensions. Usability of this solution still need to be tested with users but practical examples will complete this paper as well as a critical analysis of the preliminary results obtained.

References

Andrienko G. and Andrienko N. (2005) *Exploratory analysis of spatial and temporal data: a systematic approach*, Springer-Verlag, Berlin, 703p.

Andrienko G. and Andrienko N. (2003) Informed Spatial decisions through coordinated views, *Information Visualization*, vol 2, pp. 270-285

Bédard, Y., J. Han, 2008, Fundamentals of Spatial Data Warehousing for Geographic Knowledge Discovery, In: Geographic Data Mining and Knowledge Discovery, 2nd edition, Chap. 3, Taylor & Francis. **Accepted**

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

Béguin M. and Pumain D. (2003) *La representation des données géographiques: statistique et cartographie*, Armand-Colin, Paris, 192p.

BERTIN, J. (1973) *Sémiologie graphique les diagrammes, les réseaux, les cartes,* Deuxième édition, Éditions Gauthier- Villars, Paris, 431p.

Dyke J., MacEachren A.M., Kraak M.-J. (2005) *Exploring Geovizualization*, Elsevier Ltd, Oxford, 710p.

Fortin C., R. Rousseau(1992) *Psychologie cognitive, une approche de traitement de l'information*, Télé-université, Sainte-Foy, 434p.

Gahegan, M., Wachowicz, M., Harrower, M. and Rhyne, T.-M., 2001. The integration of geographic visualization with knowledge discovery in databases and geocomputation. *Cartography and Geographic Information Science*, 28(1): 29-44

Kraak M.-J. et Omerling F. (2003) *Cartography : Visualization of Geospatial Data*, 2nd Ed.,Pearson Education Ltd, Essex, 205 p.

Malinowski E., Zimanyi, E., 2008, Advanced Data Warehouse Design From Conventional to Spatial and Temporal Applications, 436 pp.

MacEachren A.M., 1995, *How maps work? : Representation, visualization, and design*, Guilford Press, New York, 513p.

MacEachren, A.M. (1994) *Some Truth with Maps: A Primer on Symbolization and Design*, association of American Geographers, Washington, 130p.

Newell, A. (1990), *Unified theories of cognition*. Cambridge, MA: Harvard University Press. 549 p.

Rivest, S., Y. Bédard, M.-J. Proulx, M. Nadeau, F. Hubert, and J. Pastor, 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) "Advances in spatio-temporal analysis and representation" 60 (1), pp. 17-33.

Proulx, M., Y. Bédard, 2008, Fundamental Characteristics of Spatial OLAP Technologies as Selection Criteria, Location Intelligence 2008, April 29, Santa Clara, CA, USA.

Scotch M., Parmanto B., and Monaco V., (University of Pittsburgh, USA), 2007. Usability Evaluation of the Spatial OLAP Visualization and Analysis Tool (SOVAT), *Journal of Usability Studies*, Volume 2, Issue 2, February 2007, pp. 76-95

Slocum T.A., McMaster R.B., Kessler F.C., Howard H.H., 2005, *Thematic cartography and Geographic visualization*, 2nd ed, Pearson Education Inc, Upper Saddle River, 518p.

WARE, C. (2000) *Information Visualization: perception for Design*, Academics Press, London, 438p.