GeOxygene: built on top of the expertise of the French NMA to host and share advanced GI Science research results

Bénédicte Bucher, Mickaël Brasebin, Elodie Buard, Eric Grosso, Sébastien Mustière, Julien Perret

Institut Géographique National, Laboratoire COGIT
73, avenue de Paris 94160 Saint Mandé, France
{benedicte.bucher, mickael.brasebin, elodie.buard, eric.grosso, sebastien.mustiere, julien.perret}@ign.fr

Abstract : Back to the 80ies and 90ies, researchers at the COGIT laboratory of the French National Mapping Agency often developed their own vector platforms dedicated to their respective research purposes like generalisation, integration or 3D. This multiplicity of platforms became an obstacle to many important research activities like capitalising and sharing. In 2000, based on ISO/OGC standard for data exchange, the COGIT laboratory decided to invest in the development of an ad hoc interoperable GIS platform that would host most of its research developments. This platform, called GeOxygene, was designed to have a rich data model and a genuine programming language to manipulate data. It facilitates the design and prototyping of processes which geographical data model are object oriented. It is also intended to facilitate the sharing of knowledge and programming efforts with other research labs. It is also shared through regular opensource relases. This paper highlights the relevance for a research team like COGIT to contribute to opensource software.

Introduction

The research department of the French National Mapping Agency, IGN, is composed of four laboratories. One is working on geodesy, another on hardware issues (like digital cameras), another on image processing and the last one, COGIT[1], on the management and usage of interpreted geodata. Some COGIT research activities address IGN production which include automating generalisation and label placements, or automated data matching and integration for quality controls or update process. Other COGIT research activities target external end users' needs like providing better interfaces for users to access relevant data and processes or providing better data models and methods for specific application domains.

In the 80s and 90s, using existing proprietary GIS software dedicated to vector data was an issue for COGIT research activities because these software, though they provide ‘toolboxes’ and APIs, were not completely open, i.e. their internal data models were not entirely visible. Furthermore, these software were often associated with *ad hoc* programming languages (*e.g.* Avenue language for ESRI ArcView), leading to additional training of developers and poor code re-usability. In other words, we lacked in 2000 an Open Source software to handle vector geographic data associated with an object oriented programming language (C++ or Java).

This is the reason why, at that time, COGIT researchers often developed their own platforms dedicated to their respective research purposes like generalisation, integration or 3D. Depending on their research area they would use different programming languages like ADA, C, LISP and so on. This multiplicity of platforms became an obstacle to many important research activities like capitalising and sharing geoknowledge and geo-applications. For example, as detailed in further sections, four PhD theses have been lead since 1997 on 3D modelling using different platforms. Besides, these were non-standard proprietary, independent or non interoperable platforms which did not encourage exchanges with other researchers.

In 2000, the COGIT laboratory decided to invest in the development of an *ad hoc* GIS platform that would host research developments, except for those related to generalisation which were hosted on a dedicated plateform designed in the context of a European project. This *ad hoc* GIS platform was designed to have a rich data model and a genuine programming language to manipulate data because COGIT expertise lies there: the modelling of geographical data and knowledge to improve and automate specific processes. This platform is called GeOxygene (Badard and Braun 2004) and is released under the terms of the LGPL[2] (GNU Lesser General Public License) license. It uses standard ISO/OGC models to represent geograph-

[1]. Acronym for Conception Objet et Généralisation de l'Information Topographique (http://recherche.ign.fr/labos/cogit/)

[2]. http://www.gnu.org/licenses/lgpl.html

ic information. It also uses the wide spread java language which offers interfaces to native languages. There are regular Open Source releases of GeOxygene which contain the core packages and some specific packages (basic functions, data matching, 3D, semiology, etc.) that IGN has decided to share as Open Source software, like for instance the data matching package.

The remainder of the paper is not limited to the current Open Source release of GeOxygene. We rather describe the usage of GeOxygene in our laboratory and then focus on different specific packages. Some of these packages are part of GeOxygene Open Source release.

GeOxygene: a platform to host and share research developments

GeOxygene is a development platform and not a generalist GIS platform. It is dedicated to hosting programs developed by COGIT researchers. Consequently, not much attention was initially paid to the development of GIS support functions which were not targeted by these researches, like import, export and visualisation functions. Rather, the data model has been the main focus. GeOxygene's specificity is to load geographical data into an object oriented model and to seamlessly manage the binding between a storage model and a working object model. In other words, GeOxygene facilitates the design and prototyping of processes which geographical data models are object oriented.

An important aspect of GeOxygene is the sharing and capitalizing of knowledge. Several strategies exist in order to share knowledge based on GeOxygene.

The first strategy consists in using a revision control system so that all researchers may load and see the up-to-date code of every researcher in the laboratory. To enhance code sharing based on such tools, catalogs are being developed to index all these libraries by the functions and models they support (Abd el Kader and Bucher 2006).

Another strategy is the dissemination of an open source version of GeOxygene among the community of researchers. Typically, this is a good incentive to keep effective collaboration with former COGIT researchers. COGIT's GeOxygene version is not fully synchronized with the Open Source version (hosted on SourceForge[3]). Not fully synchronized means that relevant modules are released as they reach a certain level of maturity

[3]. http://oxygene-project.sourceforge.net/

and usability. Besides, this open sourcing allows the definition of new collaborations based on GeOxygene like the GeOpenSim project funded by the French National Research Agency (ANR). GeOpenSim aims at designing an Open Source platform for the simulation of urban dynamics (Perret et al. 2009).

The last strategy is the development of interoperable Web applications. It has been targeted from the early beginning of GeOxygene's development (Badard and Braun 2004). Indeed, using ISO/OGC data models to structure geographical data ensures that input and output data can be serialized into standard XML schemas. Developing geographical Web Services is an important activity for IGN for the upcoming information infrastructures. (Balley 2004) addressed a first challenge in developing interoperable services above structured data: assisted schema transformation. Other works focus on the development of cartographic Web services to enhance the quality of maps (Bucher et al. 2007). Last, some work aims at developing Web services for integrating historical data, like old maps, into recent georeferenced frameworks (Grosso et al. 2009).

The next sections depict specific GeOxygene libraries.

GeOxygene libraries for data matching

Due to the increasing number and diversity of available geographic information, there is a growing need for the integration of different databases into a single one making explicit the relationships between them. This requires data matching tools that explicit the links between homologous objects, i.e. objects from the different databases that correspond to the same real world entity. Matching homologous objects belonging to different databases is useful in many contexts (Kilpeläinen, 2000; Hampe and Sester, 2002; Mustière and van Smaalen 2007): it facilitates propagation of updates between databases; it paves the way to the assessment of the quality of one database compared to another one taken as a reference; it allows users to make multi-criteria analyses combining information from various sources; it helps the development of efficient data visualisation tools navigating between points of views.

In this context, the COGIT laboratory has developed data matching tools specifically dedicated to matching networks with different levels of detail (Mustière and Devogele 2008). These tools are based on the comparison of geometrical and topological aspects of networks. They are in particular able to match one node of the less detailed network to a set of nodes and edges representing a complex crossroad in the other networks. Figure 1 illustrates different cases where a node of the less detailed network is matched, respectively, to a roundabout (Figure 1, 1), a shifted junction (Figure 1, 2) or a square junction (Figure 1, 3).

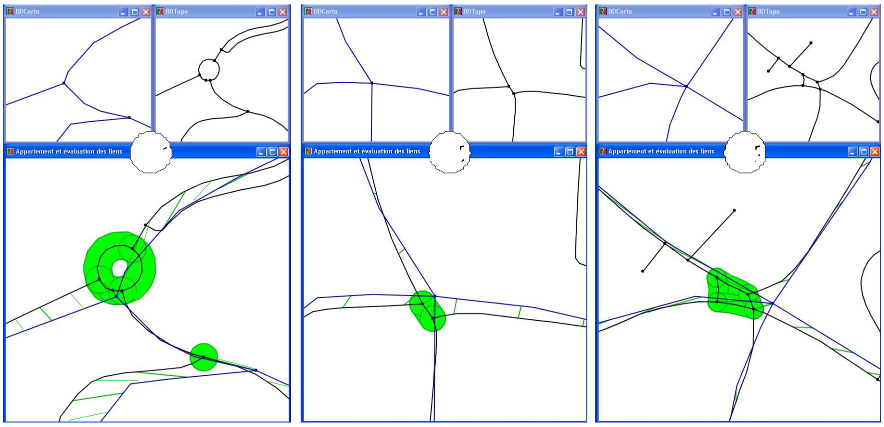

Fig. 1. Data matching results. For 1,2 and 3: the less detailed network (top left), the most detailed one (top right), and their superimposition and matching (bottom) are represented.

This tool has been released in Open Source in 2008. This concerns the data matching algorithms themselves, but also OpenJump plugins for the simultaneous visualisation of different networks, as shown in Figure 1. Compared to existing tools like OpenJump Java Conflation Suite, GeOxygene algorithms focus on the analysis of topological relationships in networks, and are especially dedicated to the matching of networks at different scales, leading to numerous one-to-many matching links. It has already been tested by several institutions such as the Belgium National Geographic Institute and the Ordnance Survey.

Other works on this subject concern a new approach for data matching based on a multi-criteria analysis of feature characteristics, including descriptive attributes. It uses techniques inspired by the so-called evidence theory to combine these criteria (Olteanu and Mustière, 2008).

Management of colour contrasts in maps

Another domain where IGN and COGIT have a solid expertise is the drawing of legible maps based on cartographic theory, including semiology of graphics. Indeed, part of IGN's vocation is the design and production of maps. Drawing legible maps is not only an issue of accessing and drawing data. Chesneau et al. (2005) aimed at providing automatic mechanisms to enhance the legibility of risk maps based on the semiology of graphics theory (Bertin 1983). This theory is a framework for the analysis of the meaning conveyed by graphical signs and provides systematic rules about how a reader sees meaningful things in maps. In other words, if there is a graphical relationship between two objects on a map, the reader will infer a specific relationship between portrayed geographical objects. For instance, if graphical objects have similar hue, the reader will infer that underlying geographical objects belong to the same theme (i.e. are of close nature). In their proposition, Chesneau et al. (2005) analyse the colours of neighbouring cartographic objects on the map and automatically enhance their colours to better render the effective relationships between them. Bad colours were found in two types of situations:

- the colours are not contrasted enough and thus not legible enough compared to the other colours,
- the colour do not respect the meaning of the corresponding theme.

The relative positions of all used colours give meaning to the themes, according to their differences. These relationships set up a comprehensive understanding of the whole map.

To facilitate colour analysis, we integrate in the process a colour space reduction obtained by creating a chromatic circle of 12 hues, each derived into 7 values. This repartition is used to visualise relationships: association and difference are represented respectively by low angle and high angle between hues, whereas order is represented by value increasing within the same hue. Following works focus on the determination of contrasts marks for each couple of colours in this chromatic circle (Buard and Ruas, 2007). These marks have been obtained through user tests where cartographic experts were asked to give scores to specific colour contrasts: hue and value. These scores are further used to compute contrast marks of hue and value between two neighbouring objects on the map.

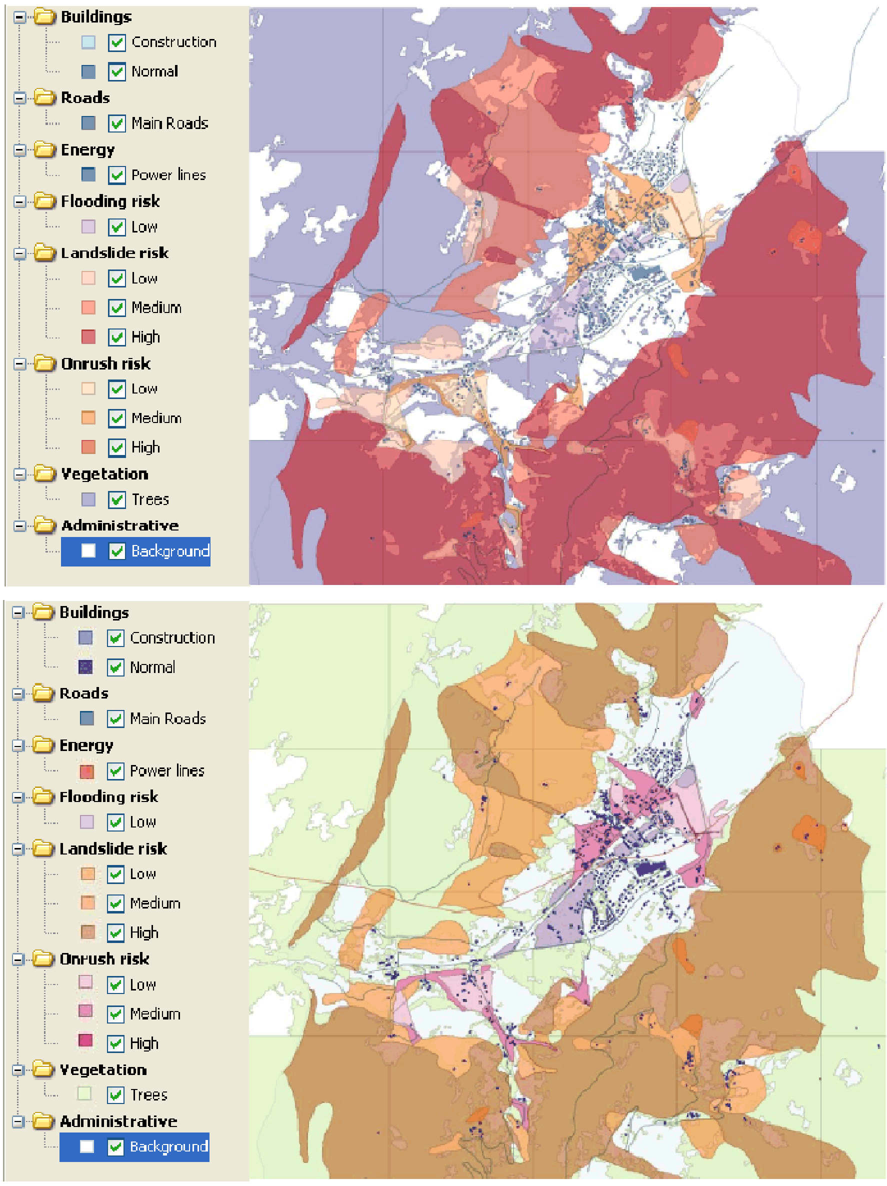

Fig. 2. The initial map (top) is automatically improved (bottom).

The module graphical user interface is designed as an OpenJump plugin. It can process any data with any colour. The module enables the user to test his own data and to be creative. It has been successfully tested on risk data which are quite complex to portray.

To activate the analysis and improvement of colour contrasts, we use two menus. The analysis menu is activated only once: it deals with identifying neighbouring objects on the map. The improvement menu can be activated as many times as wanted, knowing that each time a new state with a change of colour appears. In this way, the state cycle is not automated but supervised by the user. He can stop when he is satisfied.

The package contains Java classes supporting the definition of a cartographic structure:

- the family of features which are portrayed with the same colour (making up a theme),
- the links between neighbouring features on the map.

From these classes, the analysis of colour contrast is done first at the level of neighbouring features on the map, then for the families by aggregating the contrast marks. To improve the map legibility, the family which has the worst aggregated colour contrast mark is selected and another hue or value is proposed. New contrast marks are then computed.

Figure 2 shows an example of colour contrast improvement. The resulting map improves the rendering of theme membership through colours: two different themes, main roads and power lines, initially had the same hue; in the improved map, their colours are clearly differentiated. Likewise, the order is better portrayed in the second map and two associated themes are close in colour such as buildings-construction and buildings-normal. Moreover, the thematic meaning of colours is taken into account: purple trees made the initial map difficult to read, whereas in green the reading is immediate. The landslide risks were in red, which gives a danger meaning to the initial map, whereas this map does not aim to emphasise one of the risks.

Current work focus on the improvement of new colours proposal and several problem solving methods are being implemented and tested (Buard and Ruas, 2009).

GeOxygene 3D module

The last module depicted in this paper is the 3D module. Nowadays, 3D data are increasingly available. This can be illustrated by the production of a 3D building database at IGN (called Bati3D) or the development of the CityGML standard. However, the use of these data is often limited to visualisation in photorealistic scenes; whereas the third dimension can also bring information to solve 3D specific issues. For example, 3D data can be used to calculate intervisibility between two buildings, to represent noise pollution in 3D, to study 3D fields (like ocean temperature) or to assess the impact of an antenna. Yet, there lack 3D GIS tools for studying the territory based on 3D data and analysis methods.

For ten years already, the COGIT laboratory has studied the 3D field with the achievement of four PhD theses. Importantly, all four prototypes were implemented on different platforms.

- In the first one, De La Losa (2000) proposed a 3D topological model in order to study relationships between objects. He also discussed the representation of a 3D environment in a GIS from 2D datasets.
- Ramos (2002) designed a prototype of 3D GIS applied to the military field. It enables to calculate the shortest path on a DTM and provides intervisibility tools. As developments were led for the army, the laboratory could not keep the code, but only the description of the algorithms used during the thesis.
- Rousseaux (2003) provided methods to assess DTM errors; he demonstrated how to enhance a DTM with topographic data and showed the influence of errors in flood simulation.
- Poupeau and Bonin (2006) developed a 3D GIS prototype named Cristage. It uses crystallographic concepts to provide operations such as building simplification, symmetries detection or relationship studies.

The idea of developing a 3D module on GeOxygene was motivated by two observations. Firstly, the extendibility of the geometric schema of the platform enables the development of 3D modelling and methods. Secondly, the platform will permit to integrate these 3D works and offer the possibility to extend 2D research results to 3D.

To add 3D capacities to GeOxygene, some enhancements were made in different components.

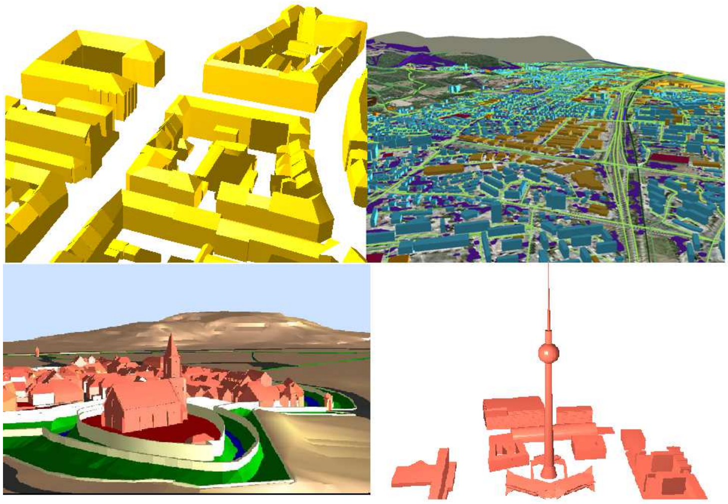

Fig. 3. Screenshots fromGeOxygene. From left to right and from top to bottom: a Bati3D dataset, ESRI shapefiles and DEM data, and 2 City GML datasets.

The first step was the improvement of the geometrical model in accordance with the ISO/OGC models (ISO 2003). Some 3D geometric classes were implemented or improved because their existence was unnecessary (for example, the GM_Solid class) or inappropriate (the GM_Surface class) in 2D. To handle 3D data, specific loaders have been implemented to handle different formats (Bati3D, CityGML, Shapefile, PostGIS).

The screenshots in Figure 3 and in Figure 4 are taken from the platform viewer. A specific interface similar to standard GIS interfaces was developed. This interface manages 3D information using a Java3D scene graph. Layer management, selection tools, light parameters and real-time navigation are provided. Comparatively a 2D standard GIS, the main difference the use of walkthrough navigation.

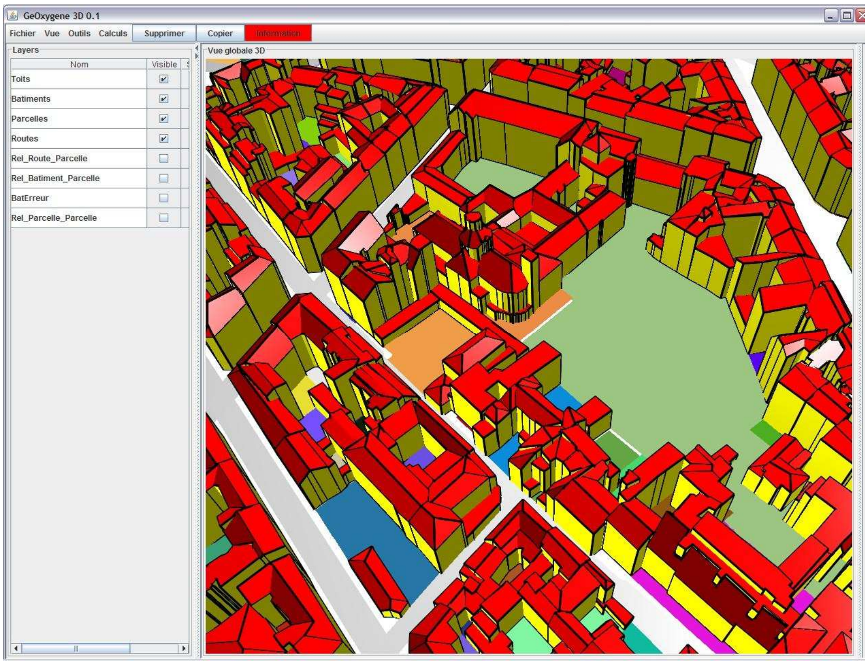

Fig. 4. Screenshot of the GeOxygene 3D interface.

3D GIS operators were developed and included in the interface. The platform provides core GIS functions in 3D such as Boolean operators (intersection, union, difference) on solids, common geometric algorithms (gravity centre, area, volume, convex hull, …) and solid decomposition (surface triangulation, tetrahedrization).

The first implementation on GeOxygene 3D module was a fast 3D simplification algorithm. It has also been developed to check the validity of the platform. 3D simplification is important in order to decrease the number of polygons in a solid. It is used to store lighter data, to represent scenes faster with fewer polygons or to communicate 3D data by Web services. Our implementation is a variant of Kada (2007). Simplification is made by merging the buffers of the different faces according to a threshold (fusion parameter). Facets of the simplified building are reconstructed by approximating buffers by the median plan. Several levels of simplification are available by changing the value of fusion parameter, as illustrated on Figure 5.

The 3D module of GeOxygene is currently in a finalization phase (the release is planned for the end of 2010). Some improvements are being led in order to ensure the robustness and the completeness of the geometrical operators (offsetting calculation, intervisibility, etc.). Some developments according to territory management are in progress; the first one deals with calculating and representing the optimal hull of buildings according to regulation constraints. Memory consumption and computation time are more important than 2D operations and can currently cause some problems. Indeed, some difficulties appear when processing on 3D geometries because the application requieres solids which are topologically valid (*i.e.* closed, no intersecting faces etc.).

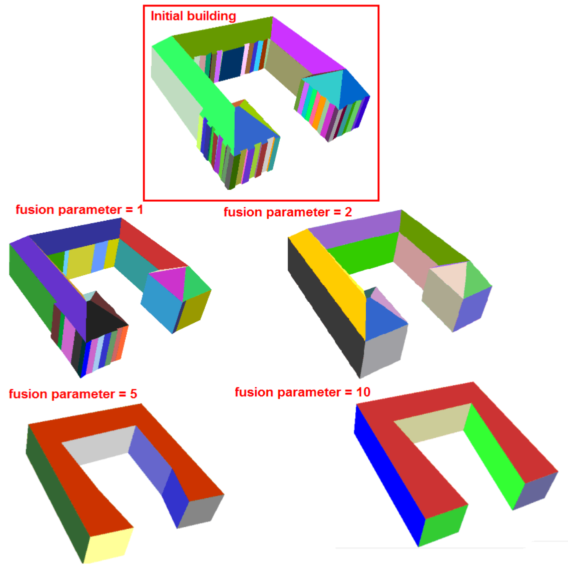

Fig. 5. Different simplification results obtained by changing the z-fusion parameter. Each face of the building is represented with a random colour to better visualize the number of faces.

Conclusion

This chapter has presented GeOxygene, a specific result and tool for COGIT research. GeOxygene's specificity is to facilitate the loading of geographical data into object oriented data models, namely into Java structures. This has allowed the development of expert modules that synthesise a long expertise in GI Science (IGN expertise) and years of research such as the data matching module and the colour contrast management module. These modules are implementations of IGN expertise. Yet, to our mind, GeOxygene is more than that. It is a central tool to support important research activities like capitalisation, collaboration, dissemination and integration. The 3D module is for example an achievement regarding capitalisation and also regarding the transfer of 2D knowledge to 3D. GeOxygene can also become a medium to transfer knowledge from IGN to the community of developers eager to develop innovative GIS applications.

References

Abd El Kader Y., Bucher B., 2006. Cataloguing GI functions provided by non web services Software resources within IGN, in proceedings of the AGILE conference, Budapest

Badard T. and Braun A., 2003, OXYGENE: An open framework for the deployment of geographic web services. Im proceedings of the 21st International Cartographic Conference (ACI/ICA), Durban, South Africa, pp 994-1004

Balley S., Bucher B., Libourel T., 2006, A service to customize the structure of a geographical dataset, Semantic Based GIS workshop (SEBGIS), Montpellier

Bertin J., 1983 Semiology of Graphics: Diagrams, Networks, Maps. University of Wisconsin Press, Madison.

Buard E., Ruas A., 2007, Evaluation of colour contrasts by means of expert knowledge for on-demand mapping, in *proceedings of the 23rd International Cartographic Conference*, Moscow

Buard E., Ruas A., 2009, Processes for improving the colours of topographic maps in the context of Map-on-Demand, in *proceedings of the 24th International Cartographic Conference, N*ovember, Santiago, Chile

Bucher B., Jolivet L., Buard E., Ruas A., 2007, The need for Web legend services, in J.M.Ware and G.E.Taylor (Eds), proceedings of the 7th International Symposium on Web and Wireless GIS (W2GIS), Springer Lecture Note in Computer Science, Cardiff, UK

Bucher, B., Jolivet, L., 2008. Acquiring service oriented descriptions of GI processing software from experts, in proceedings of the 11th AGILE Conference, Girona, Espagne

Chesneau, E., Ruas, A., Bonin, O., 2005. Colour Contrasts Analysis for a better Legibility of Graphic Signs on Risk Maps. in International Cartography Conference ICC' 2005, Proceedings, La Coruna, Spain

De La Losa A., 2000, Modélisation de la troisième dimension dans les bases de données géographiques, PhD thesis in GI Science of the University of Marne La Vallée

ISO, 2003, Geographic information — Spatial Schema ISO Draft International Standard 19107:2003.

Grosso, E., Bouju, A. & Mustière, S., 2009, Data Integration GeoService: A First Proposed Approach Using Historical Geographic Data, in James D. Carswell; A. Stewart Fotheringham & Gavin McArdle, (eds),in *proceedings of the Web* *and Wireless Geographical Information Systems*, *9th International Symposium*, W2GIS 2009, Maynooth, Ireland, December 7-8, 2009, Springer, pp. 103-119

Hampe, M., and Sester, M., 2002, Real-time integration and generalization of spatial data for mobile applications. Geowissenschaftliche Mitteilungen, Maps and the Internet, Wien, Heft(60), 167-175.

Kada M., 2007, 3D Building Generalisation by Roof Simplification and Typification. in proceedings of ICC2007, the International Cartography Association Conference, Moscow (Russia)

Kilpelaïnen, T., 2000, Maintenance of Multiple Representation Databases of Topographic Data. The Cartographic Journal, 37 (2), 101-107.

Mustière S., Devogele D. 2008. Matching networks with different levels of detail. GeoInformatica, vol.12, n.4, 12/2008, pp.435-453.

Mustière S., Van Smaalen J., 2007. Chapter 6: Database Requirements for Generalisation and Multiple Representations. in *The Generalisation of Geographic Information: Models and Applications*, Mackaness W., Ruas A., Sarjakoski T. (Eds). Elsevier. pp.113-136.

Olteanu-Raimond A.-M., Mustière S. 2008. Data matching - a matter of belief. Proceedings of the International Symposium on Spatial Data Handling SDH'2008, Montpellier, pp. 501-519.

Perret J., Boffet Mas A., Ruas A., 2009, Understanding Urban Dynamics: the use of vector topographic databases and the creation of spatio-temporal databases, in *proceedings of the 24th International Cartography Conference* (ICC'09), 15-21 November, Santiago (Chile)

Poupeau B., Bonin O., 2006, Cristage: a 3D GIS with a logical crystallographic layer to enable complex analysis, 3DGeoInfo conference, 7-8 August, Kuala Lumpur (Malaysia)

Ramos F., 2002, A multi-level approach for 3D modeling in geographical information systems, ISPRS Commission IV Symposium, 9-12 July, Ottawa (Canada)

Rousseaux F., 2003, Enrichment of a TIN with 3D vectors data, 11th Conference on GIScience and Research in UK (GISRUK'03), 9-11 April, London (UK)