

OGRS 2009

## **Geoxygene: built on top of the expertness of the French NMA to host and share advanced GI Science research results**

**Bénédicte Bucher, Mickaël Brasebin, Elodie Buard, Eric Grosso, Sébastien Mustière**

COGIT Laboratory  
Institut Géographique National  
2 avenue Pasteur  
94 160 Saint Mandé

### **Introduction**

The French National Mapping Agency, IGN, research department 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, on the management and usage of interpreted geodata. Some COGIT research addresses IGN production needs like automating generalisation and label placements, or automating data matching and integration for quality controls or update process. Other COGIT researches target external end users needs like providing better interfaces for user to access relevant data and processes or providing better data models and methods for specific domains.

In the 80ies and 90ies, using existing GIS software dedicated to vector data was an issue for COGIT researches because these software, though they provide 'tool boxes' and APIs were ultimately not totally 'open' in the sense where their internal data models were not totally visible. Developers of innovative complex programs did not have access to every part of the data model they needs. In other words, we lacked an opensource software to handle vector geodata associated with a genuine programming language.

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 they would develop in different languages like ADA, C, LISP and so on. This multi-

plicity of platforms became an obstacle to many important research activities like capitalising and sharing. For example, as detailed in further section, four PhD thesis on 3D modelling have been lead on different non interoperable platforms. Besides, these were non standard proprietary 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 most research developments. This 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 Geoxygene (Badard and Braun 2004). It uses standard ISO/OGC models to represent geographic information. It also uses the wide spread java language which offers interfaces to native languages. There are regular opensource releases of Geoxygene which contain the core packages and some specific packages that IGN has decided to share as open source software, like for instance the data matching package.

The remaining of the paper is not focused on the opensource 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 Opensource release. Others are so far exclusive to the Geoxygene COGIT version.

### **Geoxygene: a platform to host and share research developments**

Geoxygene is a development platform and not a generalist GIS platform. It is dedicated to host programs developed by COGIT researchers. Consequently, not much attention was initially paid to the development of support GIS functions which were not targeted by these researches like import, export and visualisation. Rather, the focus is on the data model. Geoxygene specificity is to load geographical data into an object oriented model and to manage the linkage between a non object storage model and a working object model, in a way seamless to the developer. In other words, Geoxygene facilitates the design and prototyping of processes which geographical data model are object oriented.

An important aspect of Geoxygene is the sharing and capitalizing of knowledge. Several strategies are possible to share knowledge based on Geoxygene.

First strategy is purely internal to COGIT: Geoxygene code is hosted on a versioning 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 tool, catalogues are developed to index all these libraries by provided functions and models (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 COGIT doctors who are working on a Post Doc in other laboratories. COGIT version is not fully synchronised with the sourceforge. Relevant modules are released from time to time when they have reached a certain level of maturity and usability. Besides, this open sourcing allows the definition of new collaborations based on Geoxygene like the GeOpenSim project funded by the Agence Nationale pour la Recherche. Geopensim aims at designing an open source platform for simulation in the urban domain.

Last strategy is the development of interoperable Web applications. It has been targeted from the early beginning of Geoxygene development (Badard and Braun 2004). Indeed, using ISO/OGC data models to structure geographical data ensure that input and output data can be serialisable into standard XML schemas. Developing geographical Web Services is an important activity for IGN in the new coming information infrastructures. A PhD thesis has addressed a first challenge in developing interoperable services that manipulate structured data: automated schema transformation (Balley 2004). Other works focus on the development of cartographic Web services to enhance the quality of maps (Bucher et al. 2007). Last, the ongoing PhD thesis of E. Grosso aims at developing Web services for integration.

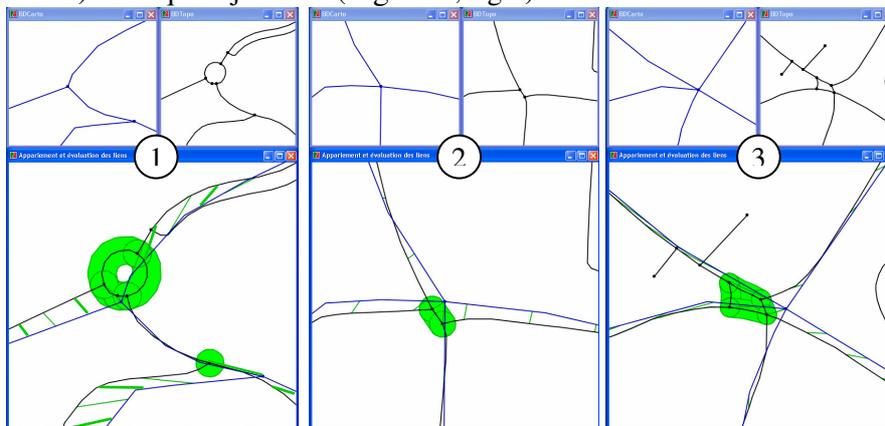
Next sections depict specific Geoxygene libraries.

### **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 re-

quires data matching tools that make explicit the links between homologous objects, i.e. objects corresponding to the same real world entity but differently represented in various databases. Identifying the links between homologous objects 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 to the development of efficient data visualisation tools navigating between points of views.

In this context, we developed in COGIT Lab data matching tools specifically dedicated to match networks with different levels of detail (Mustière and Devogele 2008). These tools are based on 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 that in different cases, where a node of the less detailed network is matched, respectively, to a roundabout (Figure 1, left), a shifted junction (Figure 1, centre) or a square junction (Figure 1, right).



**Figure 1.** Data matching results. For each figure: 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 plug-ins for the simultaneous visualisation of different networks, as shown in Figure 1. First

utilisations of it have to be noticed, like in the Belgium National Geographic Institute.

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

### **Color contrasts management in maps**

Another domain where IGN has a solid expertness is the drawing of legible maps based on cartographic theory, which includes semiology. Indeed, part of IGN vocation is the design and production of maps. Drawing legible maps is not only an issue of accessing data and drawing software. (Chesneau et al., 2005) PhD thesis aimed at providing automatic mechanisms to enhance the legibility of risk maps based on semiology theory. Semiology theory analyses the meaning conveyed by graphical signs and provides some systematic rules about how a reader see meaningful things in drawings (Bertin 1983). 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 neighbour cartographic objects on the map and change automatically the bad colours to better render the effective relationships between geographical objects. The bad colours were found in two types of situations:

- the colour was not enough contrasted and thus not enough legible compared to the other colours
- the colour did not respect the meaning of the corresponding theme.

The relative positions of all the used colours give a meaning to the themes, according to their difference. These relationships set up a comprehensive understanding of the whole map. Chesneau distinguished three kinds of relationships that can be rendered through semiology theory:

- association for close themes,
- difference for non-comparable themes
- order of importance.

To facilitate colour analysis, we integrate in the process a colour space reduction obtained by creating a chromatic circle of 12 hues, each derived

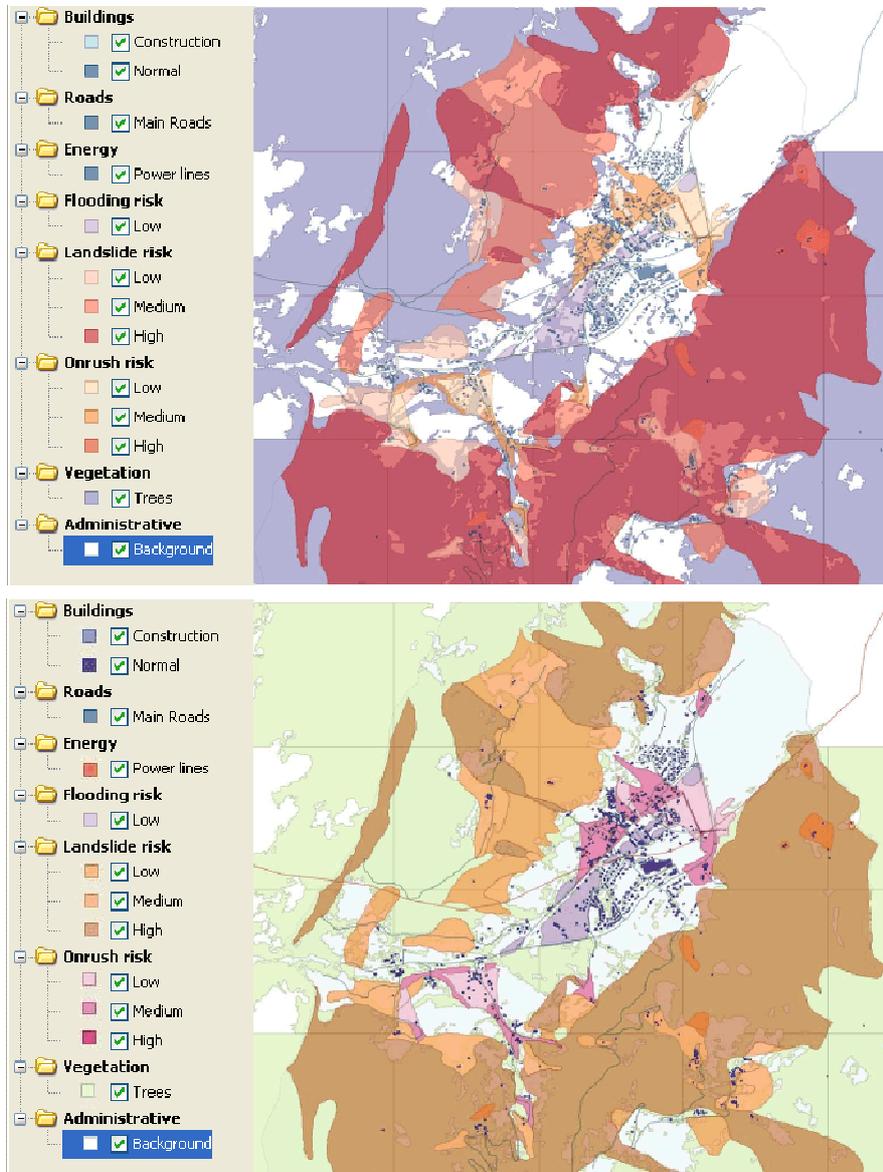
into 7 values. This repartition was necessary to visualise relationships: association and difference were represented respectively by low angle and high angle between hues, whereas order was represented by value increasing in the same hue. Following works focused on the determination of contrasts marks for each couple of colours in this chromatic circle (Buard and Ruas, 2007). These marks have been set up obtained through user tests where cartographic experts were asked to give scores to specific colour contrasts. We decided to keep the two types of contrasts: hue and value, to emphasise the relationships of the themes. Thus contrast marks of hue and value can be computed directly between two neighbour objects on the map.

The module relies on an OpenJump plugin as a graphical user interface. Any data with any colour can be used for that. It means the project does not need geographic data in the repository to run. It enables the user to test its own data and to be creative. For an application, we tested the module on risks data which are quite complex to map.

The palette of colours, the chromatic circle and the contrasts marks are hosted in a PostGIS database since they constitute the necessary input knowledge.

The core of the project concerns java classes to define the cartographic structure: cartographic objects, the family of cartographic objects that stand the colour of a theme and the creation of a class of links between neighbour cartographic objects. From these classes, the analysis of colour contrast is done first at the level of the links in reading the contrast marks, then for cartographic objects and finally for the families in aggregating the contrast marks. To improve the colours, the family having the worst colour contrast is chosen and another hue or value is proposed. A new map is displayed on OpenJump.

To activate the analysis and improvement of colour contrasts, we use two menus. The analysis menu is activated only once: it deals with looking at the neighbour 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 that way, the state cycle is not automated but supervised by the user. He can stop when he is satisfied.



**Figure 2.** The initial map (above) is automatically improved by our module. The resulting map is displayed below the initial map.

The figure above shows an example of the colour contrast improvement from the top map to the map below. The second map improves the rendering of theme membership through the colours: two different themes as

main roads and power lines had initially the same hue; they become different in the second map. Likewise, the order is better done in the second map and two associated themes are close in colour such as buildings-construction and buildings-normal. Moreover, the thematic of the colour is taken into account in the module: the trees in purple 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.

### **Geoxygene 3D module**

Last module depicted in this paper is the 3D module. Nowadays, 3D data are more and more 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 4 PhD theses. Importantly, these theses prototypes were implemented on different platforms.

- In the first one, (De La Losa 2000) proposed a 3D topological model in order to study relationship between objects. He also discussed about how to represent a 3D environment in GIS from 2D datasets.
- (Ramos 2003) designed a prototype of 3D GIS applied in the military field. It enables to calculate the shortest path on DTM and provide 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 2004) provided and implemented some methods to assess DTM errors; he demonstrated how to enhance DTM with to-

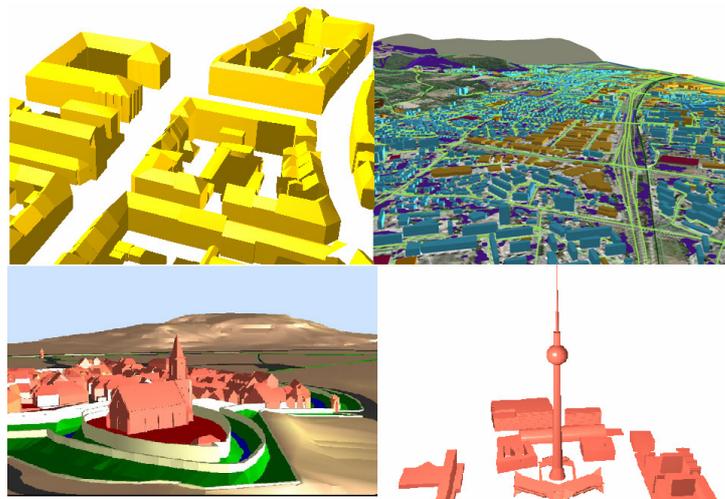
pographic data and showed the influence of errors in floods simulation.

- (Poupeau B., 2008) developed a prototype of a 3D GIS named Cristage. It was based on its own geometric model and used crystallographic concepts to provide some operations such as: building simplification, symmetries detection or relationship studies.

The idea of developing 3D on GeOxygene was motivated by 2 observations:

- firstly, the extendibility of the geometric schema of the platform enables the development of 3D modelling and methods,
- secondly, the capitalising capability of the platform will permit to integrate these 3D works and offer the possibility to extend to 3D the 2D research.

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



**Figure 3** : Screenshots from GeOxygene: from left to right from up to down: a Bati 3D dataset, Shapefile and DEM data, and 2 datasets of CityGML

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 GM\_Solid class) or inappropriate (the GM\_Surface class) in 2D. To handle 3D data, specific loaders have been added. They load data

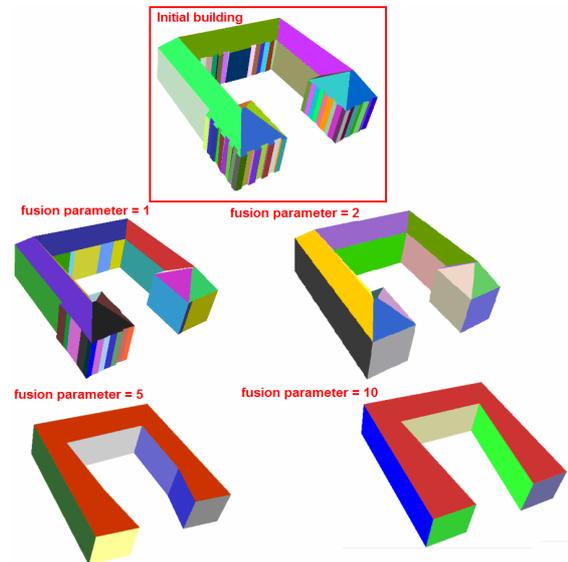
in the model from different sources (Bati3D, CityGML, Shapefile, Post-GIS).

Data on [Figure 3] are screened on the platform viewer. A specific interface similar to GIS interfaces was developed managing 3D information by the scene graph API Java3D. Layer management, selection tools, light parameters and real-time navigation are provided.

3D GIS operators were developed and included in the interface. The development needed the study and the selection of opensource external libraries. Actually, the platform is able to calculate Boolean operators (intersection, union, difference) on solids, information (gravity center, area, volume, convex hull ...) and solid decomposition (surface triangulation, tetrahedrization). Thus, necessary core of a GIS is present; it can apply the development of more advanced methods for analysis.

The first implementation on Geoxygene 3D module was a fast algorithm of 3D simplification. 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 [Figure 4].

The 3D on GeOxygen platform is currently in a phase of finalization. Some improvements are being led in order to ensure the robustness and the completeness of the geometrical operators (Offsetting calculation, inter-visibility etc.). Some developments according to territory management are in progress; the first one is about calculating and representing the optimal hull of buildings according to regulation constraints.



**Figure 4 :** Different simplifications by changing z-fusion parameter. Each face of the building is represented with a random colour to better visualize the number of faces.

## Conclusion

This article has presented Geoxygene, a specific result and tool for COGIT research. Geoxygene specificity is so far to host expert modules that synthesise a long expertness in GI Science (IGN expertness) and years of research. Typical from this are the data matching module and the color contrast manager module. These modules are implementations of IGN expertness. Yet, to our mind, Geoxygene is more than that. It is a great tool to support important research activities like capitalisation, collaboration, dissemination, integration. The 3D module is for example a great achievement regarding capitalisation and regarding also the transfer of 2D knowledge to 3D. Geoxygene is also a great tool to transfer knowledge from IGN to the community of developers eager to develop innovating GIS applications.

## Bibliography

- 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. In 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, 23rd ICA conference, Moscow
- 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*.
- 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)
- Kilpeläinen, T., 2000, Maintenance of Multiple Representation Databases of Topographic Data. *The Cartographic Journal*, 37 (2), 101-107.
- Mustière S., Devegele D. 2008. Matching networks with different levels of detail. *GeoInformatica*, vol.12, n.4, 12/2008, pp.435-453.
- Mustière Sébastien et van Smaalen John, 2007. "Chapter 6 : Database Requirements for Generalisation and Multiple Representations". Dans *The Generalisation of Geographic Information : Models and Applications*, sous la direction de William Mackaness, Anne Ruas et Tiina Sarjakoski. 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.
- Poupeau B., 2008, *Analyses et Requêtes de données géographiques 3D – Contributions de la cristallographie Géométrique*. PhD in GIScience of Université Paris Est
- Ramos, 2003, *Modelling and validating a 3D operational GIS*, PhD in GI Science of Université de Marne La Vallée
- Rousseaux, 2004, *Coherence between altimetric and planimetric data in a context of risks* PhD in GI Science of Université de Marne La Vallée