

The DOMUS Project – Building an OODB GIS for Environmental Control

Claudia Bauzer Medeiros^{1*}, Marco Antonio Casanova², Gilberto Camara³

¹ Department of Computer Science – UNICAMP
Campinas – SP – Brazil
cmbm@dcc.unicamp.br

² Rio Scientific Center - IBM Brasil
Rio de Janeiro – RJ – Brazil
casanova@riovmsc.vnet.ibm.com

³ Image Processing Division – National Institute for Space Research – INPE
Sao Jose dos Campos – SP – Brazil
gilberto@dpi.inpe.br

Abstract. DOMUS is a project of the Computer Science Department of UNICAMP, Brazil, for the development of a GIS for environmental control applications using an object oriented database. This project involves experts from different areas, both in Computer Science and in Geosciences. This paper describes the results already obtained in this project, as well as the ongoing efforts to extend available GIS technology, through use of the SPRING GIS.

1 Introduction

Commercial GIS are based on combining DBMS facilities with special purpose modules to handle specific functions, such as spatial data handling or visualization of georeferenced entities. The training for developing applications on such systems takes considerable time, since users are expected to master the different aspects of each system component, as well as the various interfaces between such components. Thus, end-users (e.g., geographers, cartographers, engineers) are seldom able to adapt existing systems to their needs, given the type of knowledge required. The DOMUS project was conceived as an attempt to help end-users solve this problem.

DOMUS is a joint research project of researchers of the Computer Science Department and the Geosciences Institute at Universidade Estadual de Campinas (UNICAMP), Brazil. Its aim is to develop GIS database tools and technology to be used by researchers in environmental control applications. In the Computer Science side, the project involves researchers in databases, computer networks, software engineering and image processing.

DOMUS' end-users (geographers and environmental planners) are expected to have some experience in using GIS, but no knowledge of DBMS technology.

* Research partially financed by grants FAPESP and CNPq

Thus, the project's goal is to provide these users with a friendly and extensible environment based on an object oriented database. This paper describes some of the results already obtained and the ongoing research and development efforts, which include using a commercial object-based GIS— the SPRING system.

The project started in 1992, and concentrates on the following issues:

- Testing the validity of using object oriented technology in the development of GIS applications, for real data.

This part of the project involves developing GIS design methodologies and tools based on object oriented databases, and checking their use against existing tools and methodologies. The advantages of object orientation have been lately advocated for this application domain [KT92], but experiments with real data are still scarce.

- Checking and extending present database research in the support of GIS requirements.

This part of the project involves analyzing the state of the art in database research for GIS environments, and proposing alternative solutions. The database research community recognizes that existing relational databases do not provide suitable platforms for GIS. Research efforts are being therefore directed towards extending existing DBMS or building new ones, in order to create such environments. However, there is urgent need for augmenting end-user participation in the development of such systems.

- Investigating the possibility of adapting or extending available GIS in order to incorporate user-tailored facilities.

This part of the project involves extending existing geoprocessing systems to incorporate new data types and procedures. GIS applications are often exploratory by nature. Unlike “standard” database applications, GIS users seldom start with an overall picture of the operations they will like to perform, and the data types to be manipulated. Thus, any GIS environment must allow user-built extensions.

- Developing database tools for GIS application design and development in the domain of environmental control.

The other three lines do not take any given application into consideration, but rather consider general requirements. This part of the project focuses on specific applications, catering to their needs. Present GIS technology, though database dependent, does not take full advantage of DBMS facilities. Thus, a database-centered approach to GIS application development is still a matter of research.

The DOMUS project has already obtained some results in the first two lines of action. Namely, an object oriented GIS design methodology has been proposed and tested [PMB93]. Also, some results on database research on object oriented views and version control mechanisms have been extended to allow manipulation of georeferenced (virtual) entities and support of temporal queries. The other two directions are now beginning to be explored, through use of the SPRING GIS [Cam93]. This software, which is now being used in more than 30 beta sites

in Brazil, has had its source code made available to DOMUS. This allows not only conducting experiments on extensibility but modifying its inner modules as well. Furthermore, it involves extending present database technology.

The remainder of this paper is organized as follows. Section 2 gives a brief overview of some related research projects. The other sections describe the research results already obtained and the present state of the project. Section 3 introduces the design methodology developed. Section 4 presents a view mechanism to allow users to manipulate georeferenced virtual entities. Section 5 discusses how DOMUS proposes to handle the problem of data evolution through use of versions. Section 6 analyzes the SPRING software, and discusses how it is going to be used to help achieving the project's goals. Finally, section 7 contains present research directions.

2 Related Projects

The domain of GIS has emerged as one of the new application areas where there is urgent need for extending database technology. Database research for GIS support has been centered on *spatial databases*, especially in what concerns algorithms for storing and accessing spatial data (e.g., [Sam89, Fra91]). Spatial structures usually support *geometric operations*, which treat geographic phenomena as if they were points, line segments and polygons (the so-called *vector* format). Related research includes geographic query languages (e.g., [RFS88, Ooi90]) and the development of GIS prototypes using new generation databases.

Several multidisciplinary research projects have recently appeared in the area. These projects involve experts not only from the database domain, but from other areas as well (e.g., computer networks, image processing, artificial intelligence). These projects can be classified in two types:

- those that are conducted within a single institution, with eventual industrial participation;
- multi-institutional projects, with industrial partners.

Projects in the first group are more numerous and may be differentiated by the nature of the data management system adopted. Thus, [HC91, vOV91] investigate extended relational architectures. [SV92, GR93, MMS93] use object oriented databases. Finally, [SRD⁺91, SA93, LL93, Web90] apply Artificial Intelligence techniques to the management of georeferenced data.

In the second group, we can quote the *Sequoia 2000* [SFGM93, Thi93] and *GOOSE* [AA93] projects in the USA, the *AMUSING* [FT93] and the *GeO2* [DRS93] projects in Europe and the *GEOTEC* effort in Brazil.

The goal of the *Sequoia 2000* project is the creation of a research and development computational environment for the area of Global Change. It involves integration of large DBMS across high speed networks, and customized visualization tools. The *GOOSE* project is developing a prototypical software platform for the management of georeferenced data, allowing communication among heterogeneous data sites. The *GeO2* project is concerned with creating an automated

cartographic environment on top of an object oriented DBMS. *AMUSING* is an ESPRIT project for defining the main features and principles of a GIS architecture, through integration of several research areas. It involves issues in data modelling, query languages, data structures and algorithms, communications and user interfaces.

GEOTEC is a multi-institutional project in Brazil to develop GIS tools and technologies. This project will start in the second semester of 1994, and involves universities, research centers and industries. Its goals are to develop prototypes and tools to solve specific problems in geoprocessing, integrating results from several domains (computer graphics, image processing, database systems, artificial intelligence, computational geometry, distributed systems and networks). Part of the DOMUS group is going to cooperate in this effort.

DOMUS is conducted inside one institution (and could therefore fit the first type of project), but it involves researchers from several fields, (which is one characteristic of the second type of project). Like all projects in the first group, it is database-centered, in the sense that methodologies, tools and techniques are treated from a database point of view. The resulting GIS environment is supported by an underlying object oriented database. In order to achieve its goals, it relies on an intensive cooperation with end-users (another characteristic of all multi-institutional projects described above). The three sections that follow briefly discuss some of the results already obtained in the last two years.

3 A Methodology for Application Modelling

One of the concerns in DOMUS is the investigation of object orientation as an adequate paradigm for GIS environments. The object oriented approach has been proposed as a means of expressing the complex relationships between geographical data and of solving the problem of diversity of formats and structures.

Object orientation supports reuse and evolution, which are needed in GIS applications. It allows other important facilities, such as the creation of new relationships and entities by aggregation of existing data, and thus isolation of users from low-level details. Furthermore, it permits modelling of the dynamic behavior of natural phenomena (using methods).

DOMUS deals mostly with data at small scales, originally in raster format. The majority of the operations performed by end users correspond to thematic overlays on raster data, and image analysis. It is, therefore, important to allow operations directly on raster data.

Object oriented GIS prototypes described in the literature use vector data for processing spatial operations. In fact, it is impractical to associate oids to raster data, unless it is first partitioned into identifiable units (e.g., by classification of cell values). Given end-users' requirements, the DOMUS project, instead, decided to accept *both* raster and vector formats.

The main points considered are [PMB93]:

- *base classes* - basic geographic features, images, and geometric features. Basic geographic features are non-spatial descriptions of geographic phenom-

ena (e.g., soil properties). Geometric features are implemented by a lattice of classes to support the vector and raster formats, with methods to handle georeferenced operations and allow combining different types of georeferenced (geometric and topologic) features.

- *primary thematic layers* – layers that are defined using a single thematic feature. In environmental control in non-settled areas, for instance, primary layers are used to describe geological structure, relief, hidrography, soil, vegetation and fauna.
- *derived layers* – layers that combine features from several sources (e.g. result of an overlay). For instance, an *erosion* (derived) layer needs to combine data on hidrography, soil, declivity and vegetation, and apply statistical and numerical methods to determine erosion hazard in given regions.

The methodology developed in DOMUS for application modelling relies heavily on end-user interaction, and is based on a class hierarchy paradigm. Most object oriented data models for GIS create georeferenced objects using only *composition* links between non-spatial and spatial objects. The DOMUS model, instead, uses *inheritance*. This is closer to the semantics of geographic applications. There is no explicit differentiation between spatial and non-spatial data. Georeferenced entities in DOMUS are treated as aggregates of other objects, some of which may have spatial characteristics, and are subject to geometric processing methods.

Derived layers are in fact an extension of materialized database views. This notion arose from the need to model spatial entities covering different spatial and non-spatial features, combining objects from base, primary and derived layers which have spatial properties of their own. The need for building derived data layers gave rise to the development of the concept of DOMUS georeferenced views.

4 Georeferenced Views

Views are usually defined as the result of a query, both in the relational model and in most proposals for object oriented systems. In DOMUS, views were found to be needed in the contexts of *restructuring* data (allowing users to see data in different ways), *integration* (to allow combining data from distinct sources and formats) and *information limitation* (selecting the desired data and hiding unnecessary details). In this context, views represent an adequate paradigm to support the following GIS requirements:

- integration of heterogeneous data sources;
- creation of derived data layers (which may be later materialized);
- display of existing data from different abstraction viewpoints;
- definition of new regions based on varying criteria (e.g. statistical, biological);
- creation of alternative scenarios for planning purposes.

The standard view definition, however, is not sufficient for GIS purposes, since users need the ability to create and manipulate *Regions*. A GIS *region* is a georeferenced entity that may portray the reality, or show imaginary scenarios

for planning purposes. Geographic regions can be specified in terms of *location (boundaries)*, *features considered* and *temporal behavior*. All these three factors can vary in the definition of a region and, at best, only “special” regions – that have fixed boundaries, with non-derived data values – are supported by DBMS. The advantage of employing a view mechanism for handling region creation lies in the possibility of better investigating users’ requirements from a database standpoint.

Georeferenced views were proposed in order to support *ad hoc region creation*. All elements that appear in a georeferenced view are considered to be *virtual* (i.e., non persistent). View data may come directly from the database, in which case it is persistent and is stored as a derived class, or originate from other views.

A *georeferenced view* extends the concept of database views. It is considered as a (virtual) database on its own, and is defined by combining a sequence of queries and special operators. The set of queries that generates the view can be applied to persistent and virtual data (thereby allowing a view to be built on other views). The special operators allow the user to perform value and/or scale transformations for all virtual entities (the so-called *generalization operations*), as well as to define boundaries, restricting the elements to be manipulated to a given spatial area (*boundary determination*). Boundary are strictly spatial operators, while generalization can apply to both spatial and thematic components. Georeferenced views differ from other view mechanisms since they explicitly require operations that must be performed before the view is made available.

Some of these views are now being simulated in O2 using its query language constructs. For instance, if **Terrain** is a set of polygonal areas containing georeferenced soil description and **Polygon1** is a predefined polygonal area, then the following view definition program allows first finding all areas with alfisol soil, and then determining only the parts of these areas that lie within the boundary determined by **Polygon1**, given a map overlay boundary operator (*BOX*). *Q* contains the result of the query (standard view definition) and the second line applies the boundary operator.

```
o2query (Q, "select t from t in Terrain where t.soil == alfisol")
for (i in Q) View += BOX (i, Polygon1)
```

5 Handling Data Evolution Management

Whereas the modelling and view generation procedures previously discussed proved adequate for handling static georeferenced entities, they did not contemplate data evolution. Temporal series (and thus temporal databases) play an important role in GIS environments, and must thus be taken into account. As stressed by [ZM92], spatial knowledge is acquired gradually. Thus, GIS application development has to evolve with data.

Furthermore, once data have been validated by the user, no direct updates are allowed. Rather, updates are performed on temporary replicates, which may or may not be transformed into permanent files subsequently. This creates the problem of maintaining different data versions.

Thus, one possibility being studied is that of using versions to allow temporal manipulation [MJ93, MJ94]. Temporal queries involve either the *what* or the *what-if* dimension. In the first case, users want to determine data validity period, or examine previous states of an entity and predict its future evolution, given its recorded behavior. In the second case, users want to perform simulation and comparison of alternative scenarios based on changing existing data.

The solution being investigated by the DOMUS project is based on the notion of database versions, and adopts the DBV version model of [CVJ94]. This model, now being simulated on the O2 database system, allows efficiently keeping track of data versions in a database. It can be embedded in any database system and does not require that the end-user control the different data configurations.

The DBV model relies on separating the notion of version as seen by the user (the *logical database versions*) from what is actually stored (*physical versioning*), and there is no replication of data, regardless of the underlying data model. This is achieved by means of an extended version stamping mechanism where the unit of logical versioning is an entire state of the universe modelled by the database (rather than just parts of the database, as in other versioning schemes).

6 Using the SPRING System

The three previous sections discussed some of the results obtained in DOMUS as regards application modelling and view and version mechanisms. This section presents an ongoing effort which involves the use of the SPRING system [Cam93]. SPRING is a commercial GIS being developed for environmental and urban applications by INPE (the Brazilian Institute for Space Research), EMBRAPA (the main agricultural research organization in Brazil) and the IBM Rio Scientific Center. At present it is under beta test conditions in several Brazilian institutions. It combines image processing, geographic analysis, digital terrain modelling functions and a spatial data base in a single environment.

SPRING presents several advantages as an experimentation platform for the DOMUS project. First, SPRING makes it possible to directly test DOMUS' object oriented methodology for application development in a commercial system. Indeed, SPRING follows an object oriented data model that provides a representation for real world geographical entities, distinguishes among their various kinds and links geometrical and non-geometrical data, uniformly accounting for vector, grid and raster data geometries. The model clearly separates the definition of high-level geographical entities from their geometrical descriptions. This permits defining very high-level GIS operations, hiding from the user the lower-level operations on data geometries. The model also features a concept similar to hypermedia links to account for the relationship between geographical entities their geometrical descriptions in maps.

Second, SPRING involves research and development in the extension of database technology, which is also part of DOMUS' goals. Thanks to this characteristic, SPRING has been developed using a modular and layered architecture based on DBMS. Therefore, it is possible to use different database systems underneath it,

by changing the interfaces between the upper-level modules and the supporting database. Thus, one of the experiments to be conducted is the coupling of the SPRING system to an object oriented DBMS. Two such systems are being considered at the moment - O_2 and OBJECTSTORE.

On a later stage, this experiment will be extended to investigate, in general, client-server architectures for geoprocessing applications. The current experience with real data indicates that adopting a heterogeneous multi-server architecture, with specialized servers, is almost unavoidable mostly because geoprocessing applications manipulate data with widely distinct characteristics. For example, vector data can be efficiently handled most of the time using a *page server*, an architecture adopted in O_2 and OBJECTSTORE. A page server, as the name implies, handles physical pages and has no knowledge of object behaviour. On the other hand, image data seems to require *object servers*, that handle entire objects and implement (some) of their methods. This follows because images in geoprocessing have a typical size of the order of 250M and involve methods, such as histogram calculations, image filters, etc. that require accessing the entire object content. However, this early assessment must be validated by prototyping and detailed performance analysis.

Third, the code has been made available to the DOMUS project. Therefore, it will be possible to make different experiments on the extensibility of the system itself, both of which are also part of DOMUS goals. This means implementing new access and storage methods tuned to the characteristics of the various data types found in Geoprocessing applications, as well as new query optimization and transaction processing strategies. Experiments with new user interface paradigms are also planned which, in particular, expand the merging of hypermedia and GIS concepts.

7 Conclusions

This paper presented the DOMUS project, a joint research effort of the Geosciences Institute and the Computer Science Department at UNICAMP, Brazil to develop GIS database tools and technology for environmental control applications using object oriented concepts.

One of the first results of this project was the definition of a methodology for GIS modelling using an object oriented framework. Other results involved investigating and proposing a version model for spatio-temporal queries, and the proposal of a georeferenced view mechanism. A side effect of the methodology was that it allowed establishing a common language for working with end-users. Indeed, users are not familiar with traditional CS practices for application development and documentation. The possibility of working within the methodology framework helped better understand users' needs.

The methodology was tested in the design of an application using data from the Cantareira Region, São Paulo State - an area of 2.000 km^2 . Available data, in raster format, was transformed into vector format and stored in the O_2 database system. The main advantage of using an object oriented system has proved to

be the facilities provided for incremental development. Geometric and image processing methods are now being coded. Further experiments will be conducted in order to improve the methodology.

The project is now beginning to use the SPRING GIS which will help achieve DOMUS' goals through the use of a new generation commercial geoprocessing product.

One issue to consider is that of creating appropriate interface primitives to let users interact directly with geographic objects. Finally, the present implementation experiments have not contemplated temporal evolution. The DBV version model will be adopted in order to test its suitability in handling this type of manipulation.

References

- [AA93] G. Alonso and A. El Abbadi. GOOSE: Geographic Object Oriented Support Environment. In *Proc. ACM/ISCA Workshop on Advances in Geographic Information Systems*, pages 38–43, 1993.
- [Cam93] G. Camara. Anatomia de Sistemas de Informacao Geográfica: Visão Atual e Perspectivas de Evolucao . In *Anais da IV Conferencia Latino-americana sobre SIG*, pages 157–183, 1993.
- [CVJ94] W. Cellary, G. Vossen, and G. Jomier. Multiversion Object Constellations: a new Approach to Support a Designer's Database Work. *To appear, Engineering with Computers*, 1994.
- [DRS93] B. David, L. Raynal, and G. Schorter. GeO2: Object Oriented Contribution for a Geographic DBMS? In *Proc. 4th International Conference Database and Expert Systems Applications*, pages 377–383, 1993.
- [Fra91] A. Frank. Properties of Geographic Data: Requirements for Spatial Access Methods . In *Proc. 2nd Symposium Spatial Database Systems*, pages 225–234. Springer Verlag Lecture Notes in Computer Science 525, 1991.
- [FT93] P. Franciosa and M. Talamo. ESPRIT Project EP6881 - AMUSING. *IEEE Data Engineering Bulletin*, pages 46–50, september 1993.
- [GR93] O. Gunther and W-F Riekert. The Design of GODOT: an Object-oriented Geographic Information System. *IEEE Data Engineering Bulletin*, pages 4–9, september 1993.
- [HC91] L. Haas and W. Cody. Exploiting Extensible DBMS in Integrated Geographic Information Systems . In *Proc. 2nd Symposium Spatial Database Systems*, pages 423–449. Springer Verlag Lecture Notes in Computer Science 525, 1991.
- [KT92] Z. Kemp and R. Thearle. Modelling Relationships in Spatial Databases . In *Proc 5th International Symposium on Spatial Data Handling*, pages 313–322, 1992. Volume 1.
- [LL93] Y. Leung and K. S. Leung. An Intelligent Expert System Shell for Knowledge-based Geographical Information Systems: 1 - the tools. *International Journal of Geographical Information Systems*, 7(1):189–200, 1993.
- [MJ93] C. B. Medeiros and G. Jomier. Managing Alternatives and Data Evolution in GIS. In *Proc. ACM/ISCA Workshop on Advances in Geographic Information Systems*, pages 34–37, 1993.

- [MJ94] C. B. Medeiros and G. Jomier. Using Versions in GIS. In *Proc. International DEXA Conference*, 1994.
- [MMS93] P. Milne, S. Milton, and J. Smith. Geographical Object-oriented Databases: a Case Study. *International Journal of Geographical Information Systems*, 7:39–56, 1993.
- [Ooi90] B. C. Ooi. *Efficient Query Processing in Geographic Information Systems*. Springer Verlag Lecture Notes in Computer Science, 1990.
- [PMB93] F. Pires, C. B. Medeiros, and A. Barros. Modelling Geographic Information Systems using an Object Oriented Framework. In *Proc XIII International Conference of the Chilean Computer Science Society*, pages 217–232, 1993.
- [RFS88] N. Roussopoulos, C. Faloutsos, and T. Sellis. An Efficient Pictorial Database System for SQL. *IEEE Transactions on Software Engineering*, SE-14(5):639–650, 1988.
- [SA93] R. Subramanian and N. Adam. The Design and Implementation of an Expert Object-Oriented Geographic Information System. In *Proc. 2nd International Conference on Information and Knowledge Management- CIKM*, pages 537–546, 1993.
- [Sam89] H. Samet. *The Design and Analysis of Spatial Data Structures*. Addison-Wesley, 1989.
- [SFGM93] M. Stonebraker, J. Frew, K. Gardels, and J. Meredith. The Sequoia 2000 Benchmark. In *Proc ACM SIGMOD Conference*, pages 2–11, 1993.
- [SRD⁺91] A. Skidmore, P. Ryan, W. Dawes, D. Short, and E. O’Loughlin. Use of an Expert System to Map Forest Soils from a Geographical Information System. *International Journal of Geographical Information Systems*, 5(4):431–446, 1991.
- [SV92] M. Scholl and A. Voisard. *Building and Object-oriented System – the Story of O2*, chapter Geographic Applications – an Experience with O2. Morgan Kaufmann, California, 1992.
- [Thi93] J. Thieman. The international directory network and connected data information system for research in the earth and space sciences. In *Proceedings ACM SIGMOD Conference*, pages 475–478, 1993.
- [vOV91] P. von Oosterom and T. Vijlbrief. Building a GIS on Top of the Open DBMS POSTGRES. In *Proc European GIS Conference*, 1991.
- [Web90] C. Webster. Rule-based Spatial Search. *International Journal of Geographical Information Systems*, 4(3):241–260, 1990.
- [ZM92] F. Zhan and D. Mark. Object-Oriented Spatial Knowledge Representation and Processing: Formalization of Core Classes and their Relationships. In *Proc 5th International Symposium on Spatial Data Handling*, pages 662–671, 1992. Volume 2.