O conteúdo do presente relatório é de única responsabilidade do(s) autor(es).
(The contents of this report are the sole responsibility of the author(s).)

Rule Application in GIS – a Case Study

Claudia Bauzer Medeiros
Geovane Cayres Magalhães

Relatório Técnico DCC–18/93

Julho de 1993

Rule Application in GIS – a Case Study

Claudia Bauzer Medeiros*   Geovane Cayres Magalhães

Abstract

Production rules in database systems have been used mostly for integrity-related issues (e.g., derived data maintenance, authority checking and constraint verification). This paper analyzes the need for using production rules in geographic information systems, for a special family of applications – utility management systems. This framework is applied to a real life large scale application – the development of an integrated database system for the maintenance and expansion of the telephone network in Brazil.

1   Introduction

Geographic Information Systems – GIS– are systems that perform data management and retrieval operations for *georeferenced data.* This term refers to data about geographic phenomena associated with their physical location (coordinates) and spatial relationships. Examples of GIS applications are urban planning, thematic and statistic mapping for natural resource management and utility facility mapping and management.

GIS have only recently benefited from database management systems. Most geographic systems are still based on a spatial data handler coupled to a sequential file manager, without any DBMS facility. Systems that provide database support are based on combining a relational DBMS with special handlers which manipulate specific aspects of georeferenced data. At present, some relational systems support spatial data structures, such

as quadtrees. However, these structures must be stored in relations in order to be accessed by the DBMS. Thus, spatial query processing becomes very complex.

Queries posed of GIS may require to know not only the present state of the data, but also prediction of future. Passive databases can support the first type of query, with considerable implementation problems when it involves spatial relationships. However, planning, engineering design and predictive analysis require management of temporal data, modelling of dynamic behavior and testing and simulation of alternative situations. For these cases, users should be provided with temporal and decision support mechanisms - for instance, in urban planning or supervisory (e.g., traffic) control. Current GIS do not provide such facilities. This paper uses the active database paradigm to solve many of these problems.

Active databases are systems that respond to events generated internally or externally to the system itself without user intervention. The active dimension is supported by triggers and production rule mechanisms, provided by the DBMS [ACC+93].

In most cases, rules are used exclusively for integrity management – either within a database or among heterogeneous systems (e.g., [BBKZ92]). Other applications include situation monitoring, user notification and process management in design applications. Recent surveys of production rules and active databases appear in [HW92, ACC+93]. Examples of present research in active databases can be found in [Cha92].

This paper discusses problems posed by GIS environments in an active database context. Production rules are shown to be relevant to perform at least three functions: supplying decision support facilities for planning and engineering design applications; helping monitoring of remotely acquired data; and supporting interface requirements. The issues presented here can contribute both to active database and GIS research:

- On the GIS side, it is shown how several required functions depend on a proper implementation of the active database paradigm. So far, research on GIS databases has centered on spatial database issues (e.g., implementation structures and indexing), query processing and data modelling. There have been no studies on application

of rules for cases such as reported here.

- On the active database side, this paper introduces a new set of problems posed by a real life complex application that uses massive amounts of distributed georeferenced data. This application concerns the implementation of a GIS database that will be used for the maintenance and expansion planning of Brazil's telephone network. It is being developed by the SAGRE project at CPqD-Telebrás[1].

This paper is organized as follows. Section 2 presents an overview of GIS requirements from a database point of view. Section 3 presents a specific GIS application domain – utility management systems – and discusses its requirements in an active database framework. Section 4 presents the types of rules that must be supported for this type of application. Finally, section 5 presents conclusions.

2 GIS database requirements

The database community has contributed to GIS research by developing data structures and algorithms in two fields: *spatial data structures* (e.g., quadtrees) which allow the efficient manipulation of objects in 2D and 3D space; and structures for supporting *geometric operations*, which model geographic phenomena as points, line segments and polygons. Applications that motivate this research are mostly in the areas of cartography and environmental planning.

Researchers consider that extensible (e.g. [HC91]) or object-oriented models [KT92] may provide a good basis for developing GIS applications, helping to solve many modelling and implementation problems posed by relational GIS. Results are recent, and there is a lack of experience with real data.

GIS data can be classified in three main categories [Ooi90]:

---

[1] Brazil's government telecommunications research center

3

- *conventional data* – traditional alphanumeric attributes, handled by conventional DBMS(e.g., the name of a city);

- *spatial data* – attributes that describe the geometry and location of geographic phenomena (e.g., the city map, and its location inside a state). Spatial data has geometrical and topological properties. Topological properties describe relationships among geographic entities (e.g., neighborhood, connectivity), and can be either explicitly stored, or derived [HT92]. Geometric data can be either in *vector* or in *raster* format.

  In *vector* format, entities are represented as points, lines and polygons. Vector data is typically processed by computational geometry algorithms. In *raster* format, data is represented in pixels, or *cells*, where each cell is associated with a set of textual values describing the points inside the cell. Different raster files are combined using planar map overlay algorithms.

- *pictorial data* – attributes that store images (e.g., photos) and are managed by image processing functions.

In some applications, category boundaries are not so clear-cut. For instance, in telephone network modelling, a particular data type – *network* – supports two categories which complement each other: the network physical layout (spatial) and the logical network connections, which determine valid communication paths (conventional data fields). Logical connections may impose a communication topology which is different from the physical layout topology.

Queries in GIS must combine and integrate different categories, e.g., relating street names (alphanumeric) to the street layout (spatial, in raster format), given neighborhood and connectivity constraints (computed using vector data). Typical GIS queries involve different types of features, or themes (e.g., vegetation, soil, hidrography). Present query optimization and indexing techniques are of little help in processing GIS queries, since users require retrieval based not only on conventional data values but also on spatial relationships among phenomena. Some of

the open issues to be considered for database support in GIS applications are:

- homogeneous handling of large amounts of textual and spatial data, which may be stored in different scales and levels of detail. There are countless open problems in this area, ranging from storage management to data modelling. Recent research is geared towards using object oriented systems, using computational geometry algorithms to compute spatial relationships (e.g., [Voi92, PMB93]).

- temporal data modelling, for processing historical sequences. Present temporal data models do not fully satisfy GIS needs, since both textual data *and* spatial relationships may change with time. Research on temporal databases has so far been restricted to nonspatial data evolution [JCG+92].

- new query optimization techniques. Again, this is an area where much remains to be done. Studies on query optimization concentrate on spatial data structures access, and usually disregard the associated textual information (e.g., [AS91, Fra91, KTS91]). Furthermore, optimization techniques do not contemplate combination of vector and raster data in a single query.

- need for adequate interface facilities (textual, pictorial). Several efforts are being conducted in this direction (e.g., [LRC+92, BM92]).

- support of supervision, planning and predictive activities. GIS are passive systems. There has been, so far, no attempt to use active databases to help planning and forecasting activities.

This paper concentrates on discussing issues concerning the last two items.

3 The SAGRE project – utility management using GIS

3.1 Project outline

The SAGRE project is an official Brazilian government R&D program for monitoring and planning the expansion of Brazil's urban telephone network, using a GIS relational DBMS.

The Brazilian telephone network at the moment serves 11 million subscribers, for an area of 8.500.000 km^2. The project started in 1991, is permanently staffed by 20 full time research personnel, and is expected to last for another 4 years.

The project is designing and implementing the core of a federated database system, which will integrate local databases at key urban centers. These databases will be used (both locally and country-wide) for the following purposes:

- SCADA – supervisory control data acquisition.

- Determine appropriate regional phone rates based on actual maintenance and usage costs.

- Coordinate scheduling of repair and maintenance of the network.

- Engineering design, supporting design decisions on determining network elements.

- Support expansion planning, given usage patterns, urban development trends and investment policies.

At present, telephone maintenance and planning activities in Brazil are executed on a local basis. The new system will allow integrating and controlling data throughout the entire country. This has required, among other things, defining standards for all levels of data manipulation - from storage structures to graphical display of georeferenced data. This is complicated by the fact that, up to now, very little exists in terms of GIS urban systems in Brazil.

3.2 Technical requirements

Network installation and maintenance involves considering equipment characteristics, personnel qualification and geographical features within an urban area. Urban phone lines in Brazil use both underground and aerial cable installations. Underground cables run inside ducts, which are protected by special electrical and pressure systems. Aerial cables are supported by poles and also require different protection and signal broadcast mechanisms.

The SAGRE database system must handle data from several sources, such as:

- equipment technical specification and prices (e.g., cable types and capacity, connector characteristics, pressure control devices);

- personnel – repair and maintenance teams (e.g, employee qualification and salaries);

- telephone rates (which vary among regions and even inside a given urban center);

- urban characteristics (geographical and socioeconomic data).

All this information is associated with its physical location, and geographical and geological features. Thus, the system must be able to support queries on small scale details (such as optimal pole placement in a given street) to country-wide managerial decisions (e.g., expansion investment priorities).

Even small-scale queries can be extremely complex. Pole placement, for instance, may require considering different types of constraints – cost, manpower needed, mechanical stress, electrical protection, communication characteristics, street topology, environmental reports and others.

4 Production rules for utility management in SAGRE

This section analyses SAGRE requirements from an active database point of view. It must be stressed that most of the rule requirements described are not yet available, given present technology limitations. The project intends to add rule processing facilities to some system modules. However, the full needs cannot yet be supported by any existing system.

The system will use (urban) georeferenced data from the following sources:

- Urban features – e.g., street network, land parcel and plot distribution, zoning characteristics, engineering structures (tunnels, bridges), traffic and pedestrian routes;

- Geographical and geological features – e.g., hidrography, topology, soil permeability;

- Telephone net data – e.g., underground duct placement, connectors, cables, poles, terminal boxes, telephone stations and centrals, and equipment (for transmission, pressurization, electrical and mechanical protection).

All these types of data are spatially connected. For instance, a cable is divided in cable sections. The database associates a given cable section with its physical location in an urban area, the protection and transmission equipment to which it is connected, and its physical characteristics – number of lines, actual telephone numbers supported, expansion potential, etc. Furthermore, remote sensing devices continuously provide the system with message traffic characteristics along this section. A section may contain a few thousand wire pairs[2], where each pair ultimately corresponds to a given telephone number. Monitoring message traffic for a given number involves following that pair along the cable network, and polling the system for signal transmission.

---

[2]Telephone signal transmission requires pairs of wires

4.1 Rules for SCADA support

SCADA is the activity of monitoring data collected from remote sensing devices in order to optimize facility usage. In SAGRE, this activity is used to monitor message traffic along telephone lines.

Most monitoring activities are presently performed by human experts who continuously examine graphic displays showing network message activity. These displays are fed with data from remote sensing data monitors that provide information about different types of network characteristics. Human controllers can act on display information to optimize message traffic flow along the net, rerouting signals. Another monitoring function is that of determining repair schedules.

Decisions are locality-dependent: an emergency situation in one urban area may have a different characteristic in a less densely populated region. Repair solutions depend on availability of repair teams and type of equipment, and geographical and physical installation conditions. Rerouting plans depend on the net topology, availability of alternate routes, and message traffic along these routes. The number of variables involved in SCADA is large, and many decisions are not taken for lack of proper support tools.

Thus, this type of activity is a prime candidate for rule support. Brazilian monitoring centers are being reconfigured so that remote sensing data will not only feed the displays alone, but also each local SAGRE database. Thus, this data will be used not only for situation monitoring, but also to establish usage patterns, thereby helping planning traffic routing and to detect areas for installation expansion. Once this data becomes available from the database, the human experts will be able to use production rule systems to help decision making. The active database itself will be able to trigger the simpler rerouting activities, liberating human experts to more complex tasks.

One of the project's aims is to establish decision support facilities in some sites, for a well defined set of situations.

4.2 Rules for decision support in planning expansion and engineering design

The main goal of the SAGRE project is to provide a basis for expansion planning and telecommunications policy making. It will allow a large economy in the maintenance and installation of lines, by integrating data in a national scale. Design alternatives, and repair and expansion schedules will be able to consider the different types of equipment and personnel available at each site, which will allow finding out the most economical solutions (given, for instance, storage and transportation costs). The system will support planning on a long and short term, and for different locality scopes (in one urban area, or in a state or the whole country).

An engineering design or installation project, on a local scale, requires the coordination of different types of design decisions: cable placement, transmission project, electrical protection project, and mechanical stress and pressure control.

The transmission project, for instance, requires considering different types of distance measurements – e.g., within ducts, between poles – to determine where to place signal boosters and remote sensor devices. Planning of electrical protection demands determining where and how to place grounding equipment. Mechanical stress studies analyze how to install equipment to support existing stresses given mechanical forces (e.g., pole location, pole support, weight distribution). In some cases, poles cannot be installed (e.g., in the middle of a thoroughfare). In such situations, aerial cables are supported by a net of *other* (non-transmission) cables, which provide the same type of function as poles. The transmission design view will see a pole at this site, whereas the mechanical project has to consider the type of support provided, and to evaluate its installation costs.

All these activities (design views) should be integrated into an engineering design activity. The designer will indicate only the key points through which cables should pass. The system will present installation alternatives, that will consider options for equipment and material to be

used, and determine staff schedule to perform the installation.

At present, a prototype is running on SAGRE that performs all of these activities using procedural code on top of the database system. However, given the variety of possible situations, it is impossible to code all alternatives into the application code. Thus, the planning of an installation (even at local – street – level) would benefit from an appropriate rule support system.

For cases where no previous installation exists (e.g., a new urban settlement) the scheduling of work may involve periods of many years. The benefits gained by a proper decision support system can be considerable, especially since policy priorities change along the years. Thus, rules should be used to reflect not only technical constraints but also guide policy making.

Another advantage in using rules is linked to the evolution of telecommunications equipment. Present applications are dependent on equipment technology characteristics. Thus, a technology change may require major recoding. Applications should be immune to these changes, which can be achieved by the use of an active system.

4.3   Symbology rules

Graphical output in GIS presents several problems. The same georeferenced element may be represented in different scales and in different contexts. Users require different output symbols for the same characteristic, where the symbol to be used is context (e.g., query) dependent.

At present, the system used by SAGRE allows symbol definition at two levels: fast symbology and full symbology. In the first case, elements are represented by simple geometric figures (e.g., a cable is a line). In the second case, users are allowed to define features such as color, size, label type. Neither option is satisfactory. Furthermore, assignment of full symbology features is very time consuming (all symbols must be entered for each case), and feature values may enter in conflict with others.

These symbology facilities are insufficient for user needs. For instance, the same physical cable is represented by different notations along

its physical path – e.g., when it changes from underground to aerial installation. Associated identification (string) tags and labels also change accordingly. Thus, one single database entity is associated with different output conventions which vary according to its physical surroundings. This means the display of database elements depends on their spatial relationships with the surrounding area. At present, this is solved by dividing the element in sub-elements according to desired display characteristics. This creates artificial entities in the database.

In a larger scale, when an urban section is examined, users may want to outline some features and ignore others. Different views of the same region can be provided by changing the display symbology.

Thus, interface design requires support of user-defined display rules in order to accomodate different requirements. Users must be able to update the rule set dynamically, thereby customizing the output. Rules must be managed by a database so that they can be easily retrieved and reused. [DF92] discuss the use of a rule system to place names in maps, for cartography applications. In SAGRE, name placement would be one of the many components to be considered by the interface support facilities.

4.4   Other demands from rule systems

Other types of rule application should also be considered. Integrity maintenance is, of course, a prime candidate. However, in this case different types of constraint have to be considered – spatial relationship constraints (e.g., [Peu88, PKC89, ES92]).

Another application would be the support of temporal modelling and querying (e.g., [SC91]).

Finally, spatial relationships are computed by special-purpose modules attached to the database system. Such functions can alternatively be attached to rules.

5 Conclusions

This paper presented a new domain for active database applications development and research – utility management systems in GIS.

The paper stresses the need for rules not only for integrity maintenance (the standard case) but in other situations as well. First, rules should guide designers in planning the design, installation and expansion of public services – in this case, telephone lines – thus providing a decision support environment for an engineering application. Second, they are shown to be a means for support of monitoring of data acquisition. Finally, their role in helping customize output is described.

Rules are applied to both traditional and spatial data. Constraint maintenance here acquires a different meaning – that of enforcing not only traditional but also spatial invariants.

The same type of case analysis can be applied to other applications of the same family – e.g., power line installation or sewage maintenance. Many other types of GIS systems that are used for planning would also benefit from active database support – e.g., environment control.

Unfortunately, the bulk and variety of data and diversity of rules needed precludes their effective use for the specific project discussed here. In order to efficiently implement rules into SAGRE, there is an urgent need for solutions in some of active database research fronts, especially improvement of algorithms for rule processing and management, situation monitoring, as well as appropriate languages for rule specification.

References

[ACC⁺93] M. Adiba, C. Collet, T. Coupaye, P. Habraken, J. Machado, H. Martin, and C. Roncancio. Trigger Systems: Different Approaches. Technical Report Aristote-SUR007, Laboratoire de Génie Informtique, IMAG, 1993.

[Aro89] S. Aronoff. *Geographic Information Systems*. WDL Publications, Canada, 1989.

[AS91]   W. Aref and H. Samet. Optimization Strategies for Spatial
           Query Processing. In *Proceedings 17th VLDB*, pages 81–90,
           1991.

[BBKZ92]   A. Buchmann, H. Branding, T. Kudrass, and J. Zimmer-
           mann. REACH − A REal-Time, ACtive and Heterogeneous
           Mediator System. *IEEE Bulletin of the TC on Data Engi-
           neering*, 15:44–47, 1992.

[BM92]   P Boursier and M Mainguenaud. Spatial Query Languages:
           Extended SQL vs Visual Languages vs Hypermaps . In
           Proc 5th International Symposium on Spatial Data Handling,
           pages 249–259, 1992. Volume 1.

[Cha92]   S. Chakravarthy. (editor) Special Issue - Active Databases.
           In *IEEE Bulletin on Data Engineering*, december 1992.

[DF92]   J. Doerschler and H. Freeman. A Rule-based System for
           Dense Map Name Placement. *Communications of the ACM*,
           35(1):68–79, 1992.

[ES92]   M. Egenhofer and J. Sharma. Topological Consistency . In
           Proc 5th International Symposium on Spatial Data Handling,
           pages 335–343, 1992. Volume 1.

[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.

[HC91]   L. Haas and W. Cody. Exploiting Extensible DBMS in Inte-
           grated Geographic Information Systems . In *Proc. 2nd Sym-
           posium Spatial Database Systems*, pages 423–449. Springer
           Verlag Lecture Notes in Computer Science 525, 1991.

[HT92]   T. Hadzilacos and N. Tryfona. A model for expressing topo-
           logical integrity constraints in geographic databases. In *Proc*

International Conference on GIS - From Space to Territory: Theories and Methods of Spatial Reasoning, number 639 in Springer Verlag LNCS, pages 251–268, 1992.

[HW92]   E. Hanson and J. Widom. An Overview of Production Rules in Database Systems. Technical Report RJ9023, IBM Almaden, 1992.

[JCG+92]   C. Jensen, J. Clifford, S. Gadia, A. Segev, and R. Snodgrass. A Glossary of Temporal Database Concepts. *ACM Sigmod Record*, 21(3):35–43, 1992.

[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.

[KTS91]   H. Kriegel, Brinkhoff T, and R. Schneider. The Combination of Spatial Access Methods and Computational Geometry in Geographic Database Systems. In *Proc 2nd Symposium Spatial Database Systems*, pages 5–22. Springer Verlag Lecture Notes in Computer Science 525, 1991.

[LRC+92]   G. Lapalme, J-M. Rousseau, S. Chapleau, M. Cormier, P. Cossette, and S. Roy. Georoute: a Geographic Information System for Transportation Applications. *Communications of the ACM*, 35(1):80–88, 1992.

[MA92]   C. B. Medeiros and M. J. Andrade. Implementing Integrity Control in Active Databases . In *Proc. 1st International Conference on Information and Knowledge Management*, pages 310–317, Baltimore, November 1992.

[Ooi90]   B. C. Ooi. *Efficient Query Processing in Geographic Information Systems*. Springer Verlag Lecture Notes in Computer Science, 1990.

[Peu88]   D. J. Peuquet. Toward the Definition and Use of Complex
           Spatial Relationships. In *Proc 3rd International Symposium
           on Spatial Data Handling*, pages 211–223, 1988.

[PKC89]   A. Pizano, A. Klinger, and A. Cardenas. Specification of Spa-
           tial Integrity Constraints in Pictorial Databases. *Computer*,
           pages 59–71, December 1989.

[PMB93]   F. Pires, C. B. Medeiros, and A. Barros. Modelling Ge-
           ographic Information Systems using an Object Oriented
           Framework. In *Proc XIII International Conference of the
           Chilean Computer Science Society*, 1993.

[SC91]   S. Y. W. Su and H. Chen. A Temporal Knowledge Represen-
           tation Model OSAM*/T and Its Query Language QQL/T. In
           Proc. 17th VLDB, pages 431–442, 1991.

[Voi92]   A. Voisard. *Bases de Donnees Geographiques: du Modele de
           Donnees a l'INterface Utilisateur.* PhD thesis, Université de
           Paris-Sud Orsay, 1992.

Relatórios Técnicos – 1992

01/92 **Applications of Finite Automata Representing Large Vocabularies,** *C. L. Lucchesi, T. Kowaltowski*

02/92 **Point Set Pattern Matching in d-Dimensions,** *P. J. de Rezende, D. T. Lee*

03/92 **On the Irrelevance of Edge Orientations on the Acyclic Directed Two Disjoint Paths Problem,** *C. L. Lucchesi, M. C. M. T. Giglio*

04/92 **A Note on Primitives for the Manipulation of General Subdivisions and the Computation of Voronoi Diagrams,** *W. Jacometti*

05/92 **An (l, u)-Transversal Theorem for Bipartite Graphs,** *C. L. Lucchesi, D. H. Younger*

06/92 **Implementing Integrity Control in Active Databases,** *C. B. Medeiros, M. J. Andrade*

07/92 **New Experimental Results For Bipartite Matching,** *J. C. Setubal*

08/92 **Maintaining Integrity Constraints across Versions in a Database,** *C. B. Medeiros, G. Jomier, W. Cellary*

09/92 **On Clique-Complete Graphs,** *C. L. Lucchesi, C. P. Mello, J. L. Szwarcfiter*

10/92 **Examples of Informal but Rigorous Correctness Proofs for Tree Traversing Algorithms,** *T. Kowaltowski*

11/92 **Debugging Aids for Statechart-Based Systems,** *V. G. S. Elias, H. Liesenberg*

12/92 **Browsing and Querying in Object-Oriented Databases,** *J. L. de Oliveira, R. de O. Anido*

Relatórios Técnicos – 1993

01/93 **Transforming Statecharts into Reactive Systems,** *Antonio G. Figueiredo Filho, Hans K. E. Liesenberg*

02/93 **The Hierarchical Ring Protocol: An Efficient Scheme for Reading Replicated Data,** *Nabor das C. Mendonça, Ricardo de O. Anido*

03/93 **Matching Algorithms for Bipartite Graphs,** *Herbert A. Baier Saip, Cláudio L. Lucchesi*

04/93 **A lexBFS Algorithm for Proper Interval Graph Recognition,** *Celina M. H. de Figueiredo, João Meidanis, Célia P. de Mello*

05/93 **Sistema Gerenciador de Processamento Cooperativo,** *Ivonne. M. Carrazana, Nelson. C. Machado, Célio. C. Guimarães*

06/93 **Implementação de um Banco de Dados Relacional Dotado de uma Interface Cooperativa,** *Nascif A. Abousalh Neto, Ariadne M. B. R. Carvalho*

07/93 **Estadogramas no Desenvolvimento de Interfaces,** *Fábio N. de Lucena, Hans K. E. Liesenberg*

08/93 **Introspection and Projection in Reasoning about Other Agents,** *Jacques Wainer*

09/93 **Codificação de Seqüências de Imagens com Quantização Vetorial,** *Carlos Antonio Reinaldo Costa, Paulo Lício de Geus*

10/93 **Minimização do Consumo de Energia em um Sistema para Aquisição de Dados Controlado por Microcomputador,** *Paulo Cesar Centoducatte, Nelson Castro Machado*