User defined spatial business rules: storage, management and implementation – a pipe network case study

Sophie Cockcroft

Department of Information Science
University of Otago. Dunedin, New Zealand
Phone: +64 3 479-8090 Fax: +64 3 479-8311
Email: scockcroft@infoscience.otago.ac.nz

ABSTRACT

The application of business rules as a means of ensuring data quality is an accepted approach in information systems development. Rules, defined by the user, are stored and manipulated by a repository or data dictionary. The repository stores the system design, including rules which result from constraints in the user's environment, and enforces these rules at runtime. The work presented here represents the application of this approach to spatial information system design using an integrated spatial software engineering tool (ISSET) with a repository at its core.

Keywords and phrases: Spatial Information Systems Development, Integrity constraints, business rules, topological relationships

1.0 INTRODUCTION

The specification of user defined rules, commonly known as 'business rules', is part of the database design process. Most database applications have certain integrity constraints that must hold with respect to the data. The simplest example of this is specifying that a data item must be of a certain data type. There are more complex integrity constraints that, for example, govern the relationships between database records. Some constraints can be specified within the database schema and automatically enforced. Others have to be checked by update programs or at data entry. These functions are expressed in a constraint language in the application.

The major challenges for data quality management in spatial databases fall into two categories: first, ensuring positional and attribute error are minimised; and second, ensuring the logical completeness of the data. The issue of quality is thoroughly covered in (Cockcroft, 1998) thus a full discussion is considered beyond the scope of this paper. However, it is known that adopting a more rigorous approach to integrity constraint management can reduce the errors described above.

This paper presents a case study of a pipe network information system developed using ISSET. It illustrates how user defined spatial rules can be built into a spatial information system and enforced at run time.

In section 2.0 the problem that this research seeks to address is elucidated. In section 3.0, as a background to the case study, a theoretical foundation of spatial integrity constraints is developed. In section 4.0 the design and development of the integrated spatial software engineering tool is briefly discussed. In section 5.0 the pipe network case study is described. This case study is based on a real pipe development project (Frey Water Engineering, 1996) with technical detail from a number of other sources (Collins, 1997; Inddex, 1998; JJValve, 1998; Ventomat, 1997). In section 6.0 the objectives of the study are re-visited, and the extent to which they are addressed is assessed.

2.0 PROBLEM DEFINITION

For a review of database integrity and consistency with respect to spatial data the reader is directed to (Cockcroft, 1996a; Cockcroft, 1996b; Cockcroft, 1996c; Cockcroft, 1997). The research presented here addresses four specific problems identified in the literature. They are as follows:

- Database schemas currently used to define spatial databases are deficient in that they do not incorporate facilities to specify business rules (Egenhofer & Al-Taha, 1992; Kim, Garza & Keskin, 1993)

- Current Spatial Information Systems/GIS do not provide a means of storing and applying spatial relationship constraints (Günther & Lamberts, 1994),

- Current Spatial Information Systems are unsatisfactory because they do not provide the facility to specify constraints on spatial relationships on the basis of attribute values (Chadwick, 1995),

- There is no means of assessing the integrity of data either in terms of attribute accuracy or spatial accuracy (Marble, 1990).

It should be noted that one of the major causes of these shortcomings is that most SIS/GIS are not part of an integrated software environment; that is, they do not have a database management system with centralised repository or CASE tool front end. From these shortcomings arise a number of objectives for the research these are as follows

- To address the problem of spatial data consistency by enforcing constraints on data entry,

- To develop a repository as the vehicle for storing and managing business rules in a spatial context and demonstrate that its deployment results in an improvement in data quality,

- To provide the facility for specifying spatial relationships on the basis of attribute data and enforcing constraints relating to this attribute data,

- To allow for semantic information to be attached to topological relationships.

3.0 TAXONOMY OF SPATIAL INTEGRITY CONSTRAINTS

For a full discussion of integrity constraints as they relate to spatial data see (Cockcroft, 1997). Figure 1 presents one view of integrity constraints in traditional DBMS. It shows how they relate to the data model, the data definition language and the application that uses the database. The constraints applicable at each of these levels of spatial information systems development are inherent constraints, implicit constraints and explicit constraints respectively.
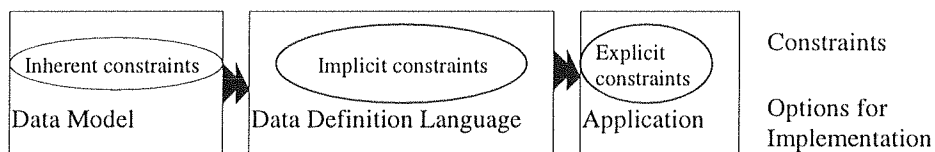

Figure 1 Classes of integrity constraint in terms of where they are applied
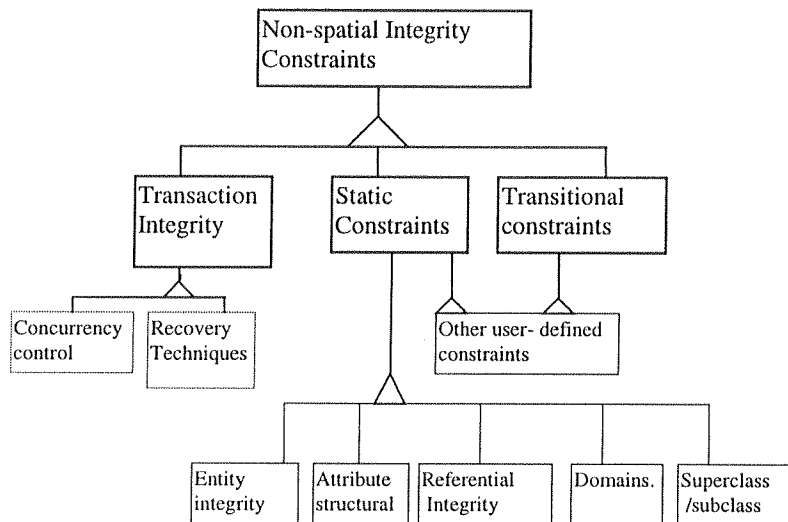
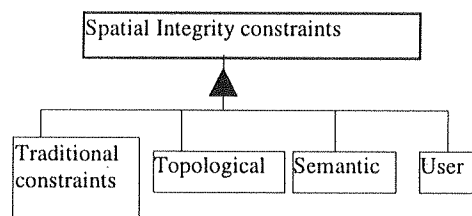Figure 2 Traditional Integrity Constraints adapted from Date (Date, 1990)

Figure 3 Spatial Integrity Constraints

Topological integrity constraints

Topology is the mathematical procedure for defining spatial relationships between points lines and polygons. There has been some theoretical research into the principles of formally defining these relationships (Egenhofer & Franzosa, 1991). The issue of defining topological integrity constraints in databases according to these principles has also been investigated (Hadzilacos & Tryfona, 1992).

Semantic integrity constraints

These differ from topological integrity constraints in that they are concerned with the *meaning* of geographical features. An often-quoted (and encountered) data quality problem is that of road centrelines not meeting. The concern is the *topological* consistency of the line object road centre line, which has implications for analysis. This could be addressed regardless of the semantic information that this is a road. Semantic integrity constraints apply to database states that are valid by virtue of properties of objects that need to be stored.

User defined integrity constraints

These differ from semantic integrity constraints that are more esoteric in nature and not necessarily based on semantics. User defined integrity constraints allow database consistency to be maintained according to *user defined* constraints analogous to business rules in non-spatial DBMS

Examples

1. **Static semantic:** The height of mountain may not be negative.
2. **Static topological:** All polygons must close
3. **Static user:** All streets wider than seven metres must be classified as highways.
4. **Transition semantic:** The height of a mountain may not decrease.
5. **Transition topological:** If a new line or lines are added making a new polygon the polygon and line tables must be updated to reflect this.
6. **Transition user:** A road of any type may not be extended into body of water of any type

4.0   DESIGN OF REPOSITORY SYSTEM

The repository stores metadata on entities and their relationships. Every entity may have one or more attibutes. There are rules that apply to attributes and also to relationships. Metadata is also stored about the geometry of entities and their graphical representation. It is suggested that assigning a graphical representation to an object at this level also helps to preserve integrity since it is easier for the user to see when they have entered an object wrongly. Finally, every object belongs to a project which means it is possible to restrict the objects a user is working on to those in the domain of interest. In terms of metadata relating to format and lineage the name of the person who entered the data and its origin is also stored, with the name of the coordinate system, coordinate units, the area units used, bounding coordinates, scale and date input.

The repository is implemented in Borland Delphi and interfaces with MSAccess and MapInfo. It should be noted that the repository is independent of the spatial information system software used and could potentially be coupled with any other SIS software. Thus it will be appreciated that the system developed is designed to illustrate a principle rather than to solve a particular problem e.g. that of pipe network design.

5.0   PIPE NETWORK CASE STUDY

This section illustrates how the repository system developed addresses the objectives outlined in section 2.0. The pipe network case study is based on a real water supply development project (Frey Water Engineering, 1996) with technical detail from a number of other sources (Collins, 1997; Inddex, 1998; JJValve, 1998; Ventomat, 1997). In order to illustrate all the types of rules discussed in section 3.0 that are of relevance to this project, strict adherence to the network plan described in project (Frey Water Engineering, 1996) has been sacrificed in places.

Referring to the work of Chadwick (Chadwick, 1995) it is desirable that constraints on spatial relationships based on attribute data should be defined and applied in a seamless fashion although both spatial and attribute data are involved. This is not a trivial process since topological relationships depend on spatial coincidence and must therefore be administered by a system which is spatially capable. Attribute information on the other hand does not require such processing and is often not stored with topology, it is stored separately and linked to spatial data through a unique identifier. The process for achieving this is becoming increasingly sophisticated thus whilst the currency of linkages between spatial and non spatial data was once a key issue in spatial data integrity it is now less so. In order to illustrate the capabilities of the repository system three sample constraints have been selected:

1.   Pipe material must be ductile iron, steel, concrete or cement
2.   Butterfly valves can only be connected to water pipes > 14" in diameter
3.   Pipes can only be connected to valves, fittings, pumps and reservoirs

The first constraint according to the taxonomy given in section 3.0 is a *static semantic* constraint, it is also an implicit rule and is implemented using domains. In the repository system developed here, the user specifies the domain from which attribute values can come and this is then included in the database schema which is automatically generated. The second constraint is a spatial constraint based on attribute data it is a *semantic rule* according to the taxonomy in section 3.0. If the spatial and attribute data were stored under the same architecture it would also be of the *implicit* type because it could be specified in the data definition language (see for example the spatially extended entity relationship schemas described in (Firns, 1994)). In fact, although the process is seamless to the user, the topological aspect of this rule is managed in the repository and passed to the SIS. In the repository the user can specify that certain spatial features

•   may not be topologically related or

•   may be topologically related according to certain conditions including conditions on attribute data

The third constraint can only be implemented where semantic information is attached to a spatial feature. It is again a *static semantic* constraint, but this time it is explicit because it is beyond the scope of the Data Definition Language (DDL). That is, code would need to be written to enforce this constraint which would take the form of triggers. In the system developed here it is possible to specify how a spatial object relates to all other spatial object within the domain of interest this will be illustrated below. In the following section it will be shown how the repository addresses these particular rules, each will be taken in turn.

In all cases the user selects the pipe network project at the opening screen as shown in Figure 4.
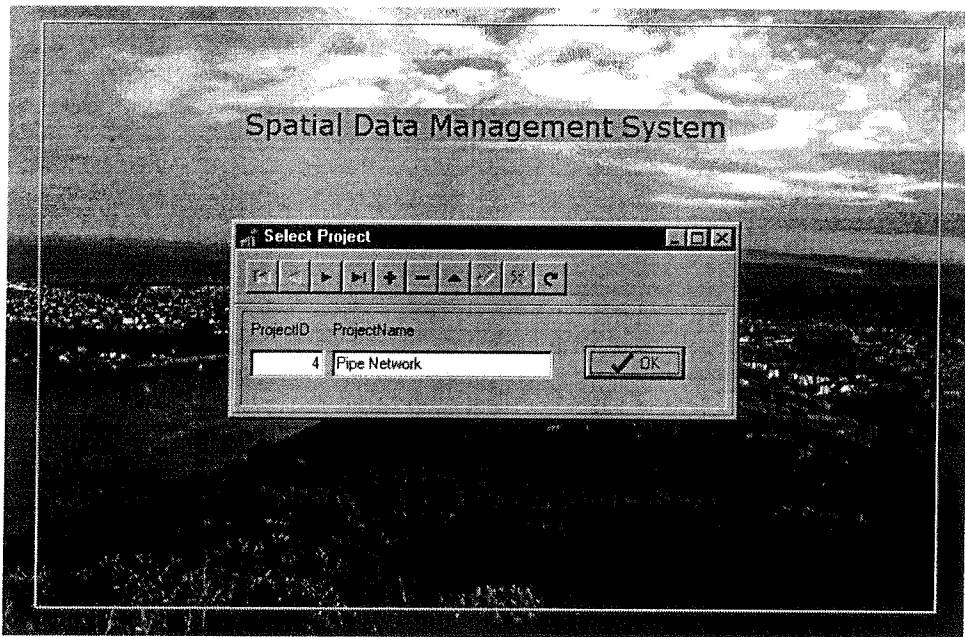

Figure 4 Opening screen of repository system

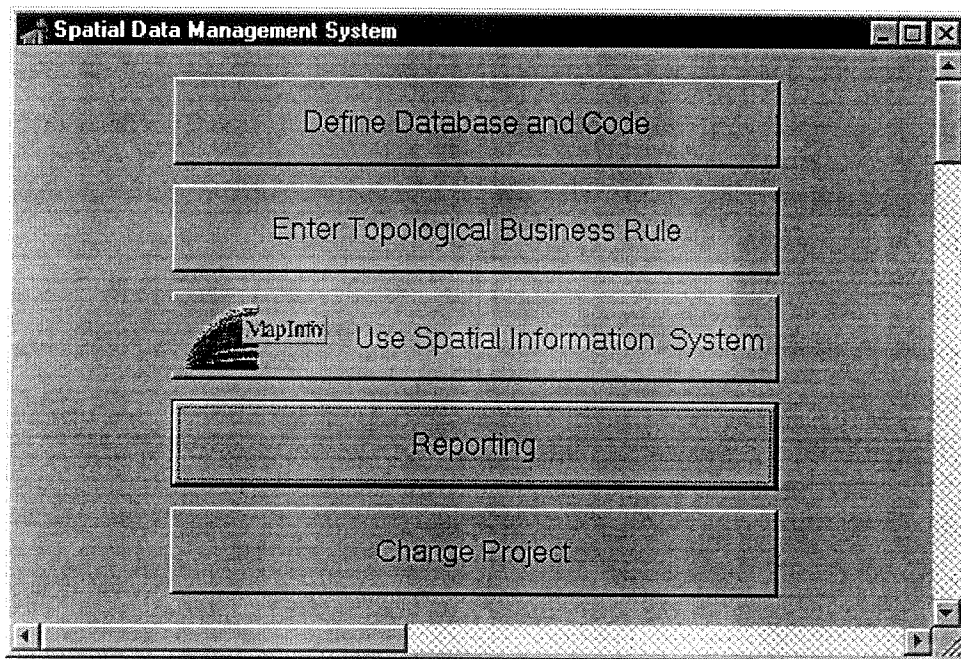## 5.1   Pipe material must be ductile iron, steel, concrete or cement

Figure 5 Repository Main Menu

The user selects define database and code as shown in Figure 5. The database can then be defined using Enter/Edit metadata as shown in Figure 6. Details of specifying allowed values for an attribute are given in Figure 7.
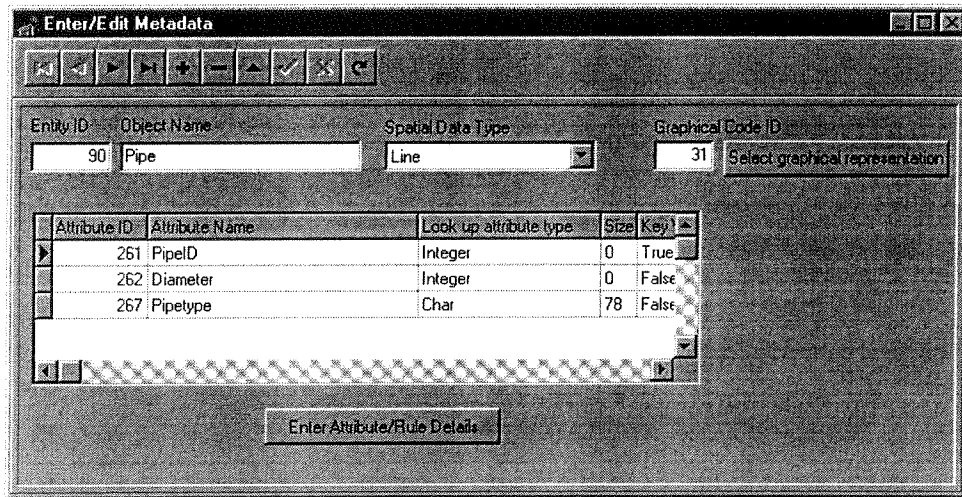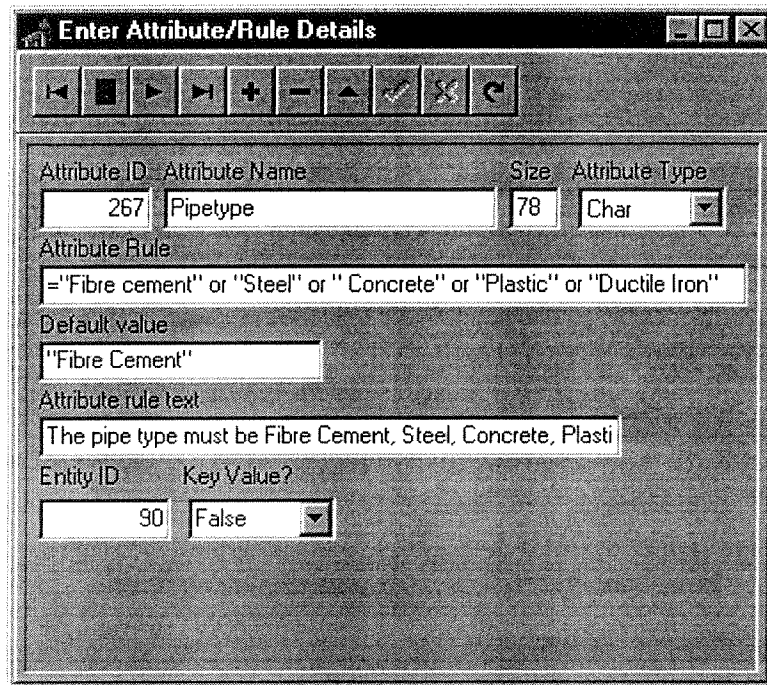
Figure 6 Metadata Entry

Figure 7 Specification of Domains

The spatial information system is then run and, once the user has entered the object, in this case a pipe, dialog boxes prompt the user for data entry. If the data entered was not from the range specified in Figure 7, the error message shown in Figure 8 displays.
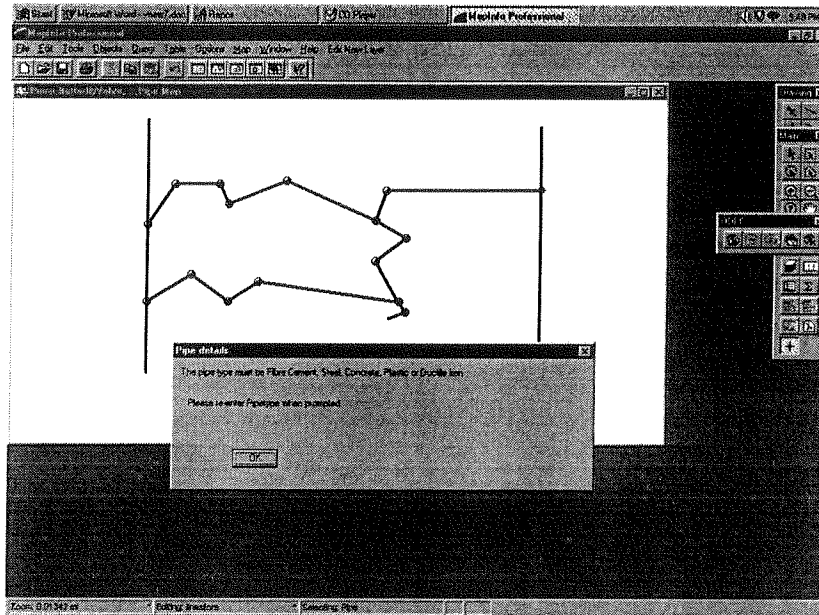
Figure 8 Domain error illustrated in pipe network case study

5.2   Butterfly valves can only be connected to water pipes > 14" in diameter

The following screen shots illustrate the process of specifying this constraint. Initially the user selects *Enter Topological Business Rule* from the screen illustrated in Figure 5.

The user is then prompted with the screen illustrated in

Figure *10*. The relationship between two entities may have one or more conditions attached as shown in Figure 9
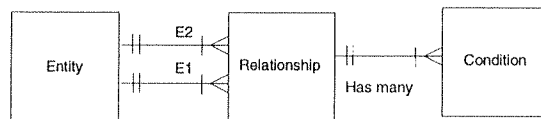

Figure 9 A single relationship between two entities can have many conditions attached to it
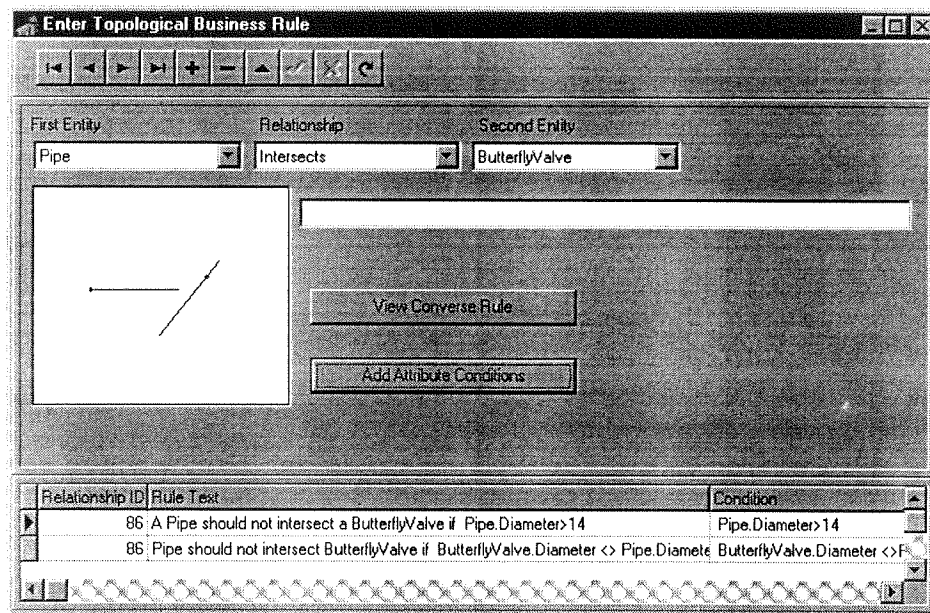

Figure 10 Entering topological constraints based on attribute values

Figure 10 shows the detail of entering a condition on the relationship. The user selects the *build expression* button to take them to the condition builder shown in and Figure 11.
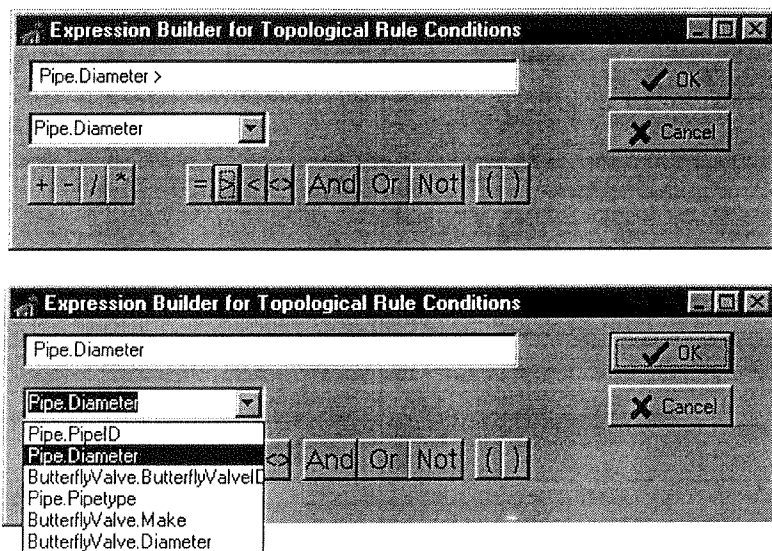

Figure 11 Attribute condition data entry screen

Turning to Figure 12 the user is then prompted to accept the converse rule. If they accept, all its conditions are also copied across, but they can be edited later.
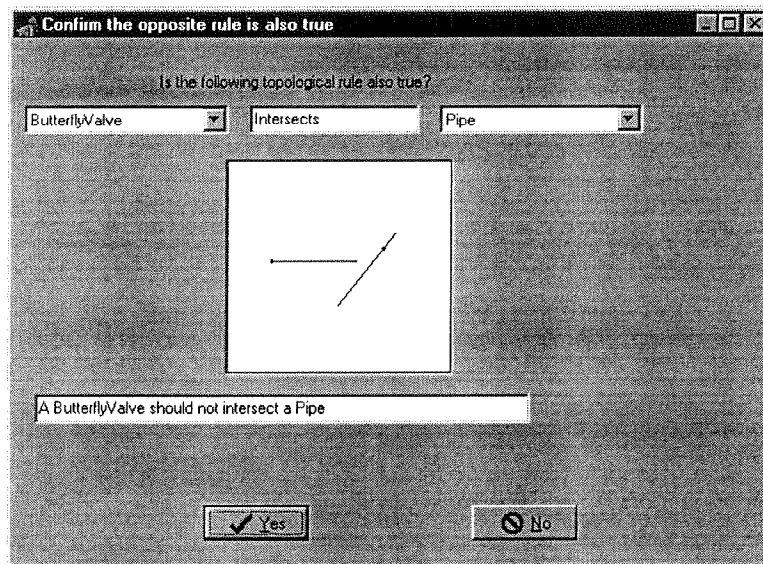

Figure 12 The system prompting for acceptance of opposite rule

It will be appreciated that there are some topological rules which are true on one side but not necessarily on the other for example a valve must intersect a pipe, but a pipe does not have to intersect a valve.
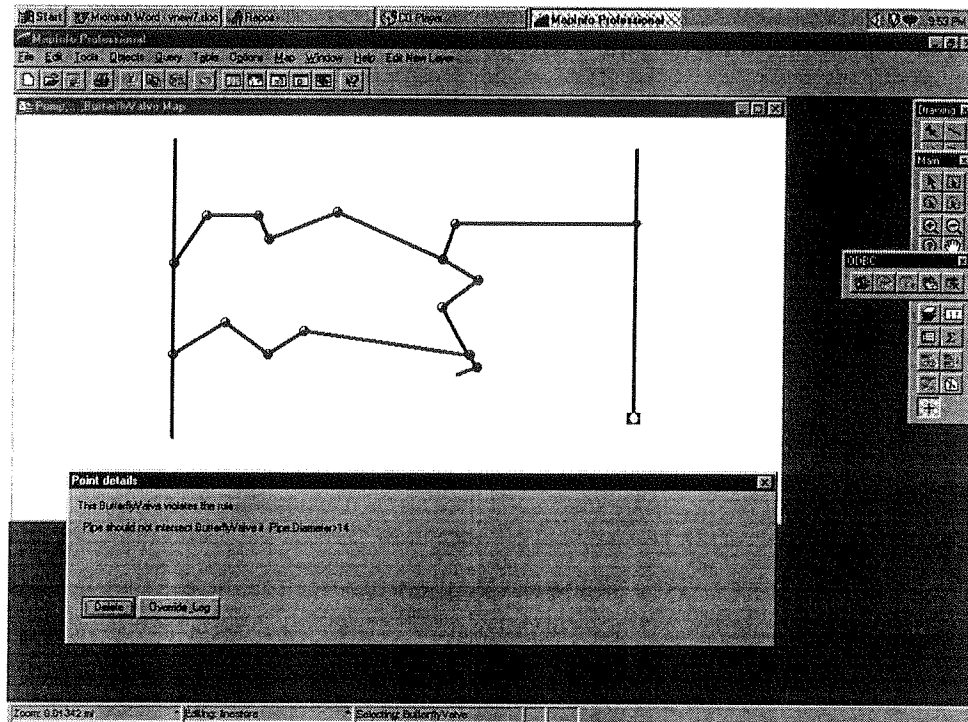
Figure 13 Error message resulting from violation of constraint 2

5.3 Pipes can only be connected to valves, fittings, pumps and reservoirs

In order to specify the third rule outlined in 5.0 the relationship between the pipe and all other objects in the project would be defined as illustrated in Figure 14. It will be observed that there is an option to override and log violations of semantic constraints that have a spatial component. In the following section, the reports generated from such a log are illustrated in section 5.4.
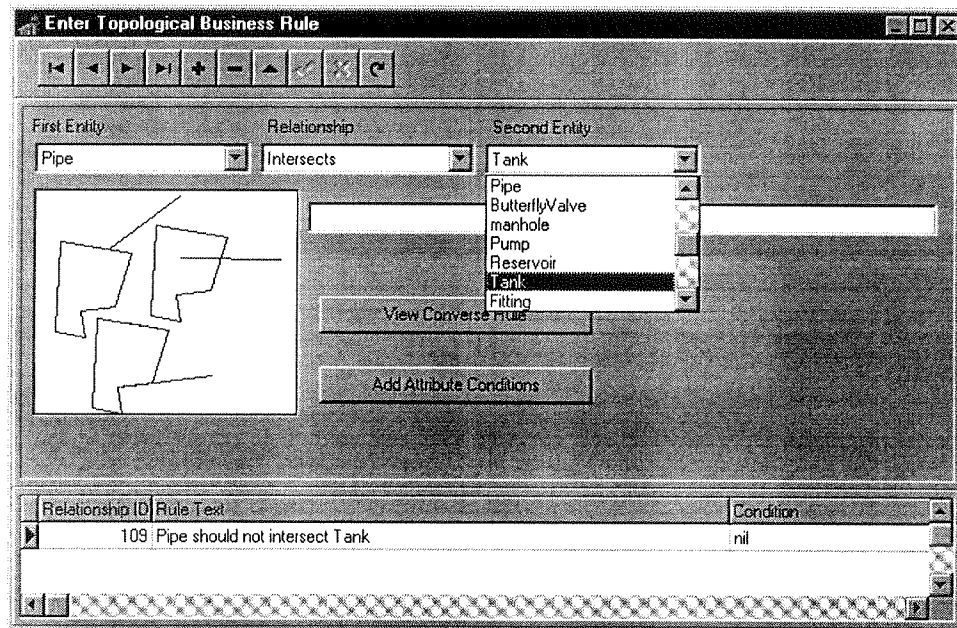

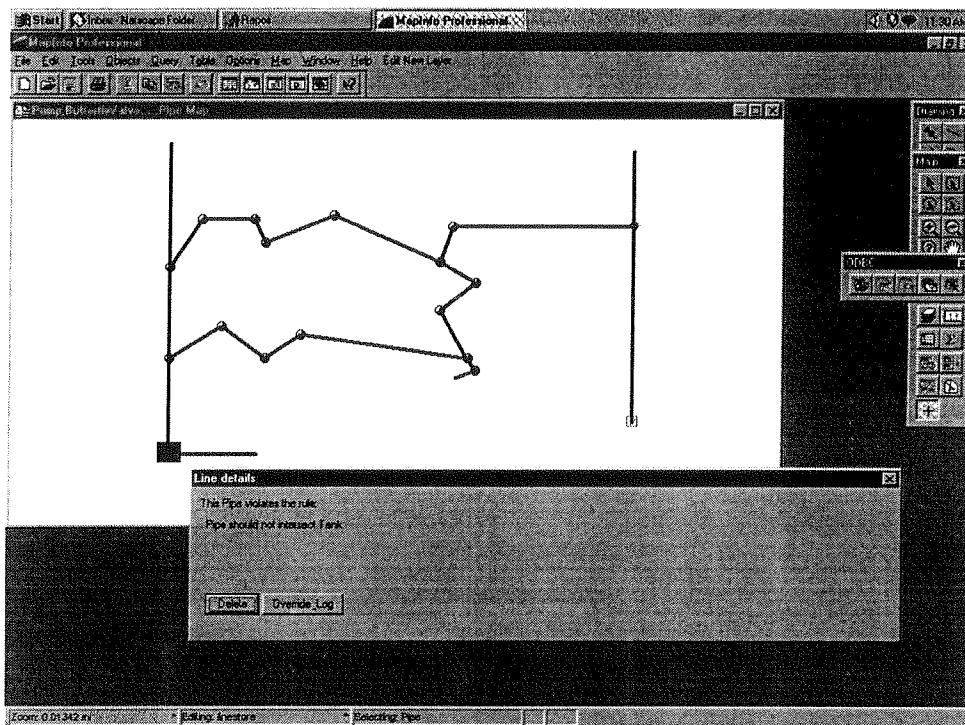*Figure 14 Storing information on illegal relationships in the repository*

Figure 15 Violation of spatial domain error

5.4 Reporting

A number of reports are produced by the repository, including a report corresponding to the FGDC standards for quality metadata (FGDC, 1994), and a full report of entities and attributes. Illustrated here is the Error Log report described in section 5.3.

1

Error Log

MAPINFO_ID	Rule	ObjectName	DateAdded	Keyvalue	MinX	MaxX	MinY	MaxY
124	ButterflyValve should not intersect Pipe if Pipe.Diameter>14	ButterflyValve	6/29/98	2	1.8e-005	-2.7e-005	1.8e-005	-2.7e-005
125	ButterflyValve should not intersect Pipe if Pipe.Diameter>14	ButterflyValve	6/29/98	4	-8.6e-005	1.3e-005	-8.6e-005	1.3e-005
126	A Pipe should not intersect a ButterflyValve if Pipe.Diameter>14	Pipe	6/29/98	5	-7.4e-005	-5e-006	-7e-005	2.4e-005
127	Reservoir should not intersect ButterflyValve	Reservoir	6/29/98	6	-9.2e-005	1e-005	-8.4e-005	1.8e-005
128	Reservoir should not intersect ButterflyValve	Reservoir	6/29/98	7	1.2e-005	-3.2e-005	2.3e-005	-2e-005

Figure 16 Error log report

7.0 RESULTS

Referring to the objectives outlined in section 2.0, the majority of them have been addressed by building the system. In order to satisfy the objective of demonstrating that deployment of the repository results in an improvement in data quality, three approaches are put forward.

1. Proof of concept; the system demonstrably *does not allow* incorrect data to be entered, so the data quality must be improved. In this case all that is needed to prove this is to show the system restricting data entry

2. Where over-rides are allowed a measure of data quality can be obtained by calculating the following automatically in the repository

$$\frac{Total\ number\ of\ Errors}{Total\ number\ of\ Relationships}$$

and

$$\frac{Total\ number\ of\ Attribute\ Errors}{Total\ number\ of\ Attributes}$$

It is desirable to allow a user to override constraints to allow for exceptional cases. By producing a count of overrides a percentage quality measure can be automatically generated. The production of these figures in itself is an improvement in quality because many spatial data sets are digitised without recourse to any error counting system, so the resultant quality is unknown.

3. A case study can be conducted with and without the imposition of constraints and allow a human user can then be allowed to peruse the resultant map.

This is not illustrated here because it is the most subjective measure, in that it depends on the opinion of a human expert, and is considered to be of limited value in this context.

8.0 CONCLUSION

This paper has illustrated the use of a repository to enforce integrity constraints in a spatial information system. These are enforced both on the attribute data and on spatial relationships between entities. The research presented here addresses a need, identified in the literature, for semantic information to be attached to spatial relationships, and for a rules system to be put into place based on these semantics.

9.0 REFERENCES

Chadwick, A. (1995). "An Architecture to support business rules in custom GIS applications". In *Proceedings: 18th AM/FM International Conference*, 1047-1058.

Cockcroft, S. (1996a). "First Experiences in Implementing a Spatial Metadata Repository". In *Proceedings: 1st International Conference on Geocomputation*, Leeds UK: R. J. Abrahart (Ed.), University of Leeds. 111-120.

Cockcroft, S. (1996b). "Towards The Automatic Enforcement of Integrity Rules in Spatial Database Systems". In *Proceedings: SIRC'96: Proceedings of the 8th Annual Colloquium of the Spatial Information Research Centre*, University of Otago: R. T. Pascoe, N. C. Sutherland, & P. Gorman (Ed.), University of Otago. 33-42.

Cockcroft, S. (1996c). *The use of a metadata repository in spatial database development* (96/02). Otago University Working Paper Series.

Cockcroft, S. (1997). "A Taxonomy of Spatial Data Integrity Constraints", *Geoinformatica*, 1(4), 327-343.

Cockcroft, S. (1998). "Spatial data quality control through the use of an active repository". In *Proceedings: Third International Conference on GeoComputation*, Bristol, UK:

Collins. (1997). *Collins Pipe & Supply Co., Inc. Provider of Pipe, Valves and Fittings* http://www.collinspipe.com/inside.htm , Accessed July 31st 1998

Date, C. (1990). "A contribution to the study of database integrity". In *Relational Database Writings 1985-1989*, (pp. 185-215), Saratoga: Addison Wesley.

Egenhofer, M. J., & Al-Taha, K. A. (1992). "Reasoning about gradual changes of topological relationships". In *GIS, From space to territory, theories and methods of spatio-temporal reasoning in geographic space, Lecture notes in Computer Science,* (vol. 639, pp. 196-219), A. Frank, I. Campari, & U. Formentini (Ed.), New York: Springer Verlag.

Egenhofer, M. J., & Franzosa, R. D. (1991). "Point-set Topological Relations", *International journal of geographical Information Systems*, 5(2), 161-174.

FGDC (1994). *Content standards for digital Geospatial Metadata* . Federal Geographic Data Committee.

Firns, P. G. (1994). "An Extended Entity Relationship Model Applicable to the design of Spatially Referenced Databases", *Ph.D. Thesis*, University of Otago, Dunedin, New Zealand.

Frey Water Engineering, I. (1996). *Seaford Rise Development, Zone 126 GA Optimization Study* http://www.mcs.com/~jfrey/seafd126.htm , Frey Water Engineering, Inc. Accessed 31 July, 1998.

Günther, O., & Lamberts, J. (1994). "Object-oriented Techniques for the Management of Geographic and Environmental Data", *The Computer Journal*, 37(1), 16-25.

Hadzilacos, T., & Tryfona, N. (1992). " A model for expressing topological integrity constraints in geographic databases". In *Proceedings: Theories and Methods of Spatio-Temporal Reasoning in Geographic Space. International Conference GIS - From Space to Territory.*, Pisa , Italy: A. U. Frank, I. Campari, & U. Formentini (Ed.), 252-68.

Inddex. (1998). *Industrial Equipment Search* http://www.inddex.com/english/home.htm. Accessed 31 July, 1998.

JJValve. (1998). http://www.jjvalve.com/frame.html. Accessed 31 July, 1998.

Kim, W., Garza, J., & Keskin, A. (1993). "Spatial data management in database systems: research directions". In *Proceedings: Advances in Spatial Databases. Third International Symposium, SSD '93*, Singapore: D. Abel, & B. C. Ooi (Ed.), Springer-Verlag Berlin, Germany. 1-13.

Marble, D. F. (1990). "The extended data dictionary: A critical element in building viable spatial databases". In *Proceedings: 11th annual ESRI user conference*

Ventomat. (1997). *AIR VALVE TECHNOLOGY REVIEWED* http://www.ventomat.com/chap0.htm. Vent-O-Mat. Accessed 31 July, 1998.