Sharing Feature Based Geographic Information - A Data Model Perspective

Joseph Greenwood and Glen Hart

Research and Innovation, Ordnance Survey,
Romsey Road, Southampton, SO16 4GU.
Telephone: +44 (0)23 8079 2125,
Email: jgreenwood@ordsvy.gov.uk, ghart@ordsvy.gov.uk

Abstract

This paper explores the role of the "Feature" as the atomic unit of geographical information exchange. It proposes the disaggregation of the feature into components so that information related to the geographical entity represented by the feature may be exchanged at a finer level of detail and examines the implications of this with respect to the geodata model.

1 Introduction

The ability to effectively share geographic information between different technical disciplines and organisations is becoming increasingly important, be this the inter-disciplinary nature of scientific research, the need for 'joined-up' governmental and commercial processes or to create mobile services for multiple consumer groups. Each domain has its own 'world view' which describes the conceptualisation of the real world objects or phenomena under consideration in the domain. From a data perspective this is often encapsulated in data capture and maintenance specifications and regimes. As Burrough and Masser (1998) note however, ".. related sciences and technical methods do not necessarily describe or recognise the same spatial phenomena in the same way". Setting aside the political or economic barriers, the technical task for sharing geographic information becomes that of being able to explicitly model data that represents differences in definition and interpretation of the same geographic phenomena. This must be done such that it enables the similarity and differences to be explicitly managed to enable human or computer decisions of whether the information can be exchanged.

To successfully share data a recipient must understand what a supplier is sending, how it differs from what is needed and how it can be translated if it does not exactly match the end user requirements (or what the impacts are if translation is imperfect). This situation is further complicated if multiple suppliers are involved. The problem can therefore be seen at various levels. There are semantic issues concerned with the nature of what is being supplied: the recipient needs building information, is that what the supplier means by "roofed structure"?. There are issues concerned with correlating information received with the features the recipient is interested in: Does this building information describe the building identified in the recipients' database as "build 2". There are issues concerning the form of the data supplied: Are the units of measurement those that are required? Does the time period for which the data is valid meet the required needs? Is the data supplied at a required level of precision? Is the data packaged in a convenient way?

At present, answering these questions is a largely manual task and often imperfect in providing a complete solution. Final implementations are typically hardwired solutions. Significant amounts of effort are often spent reconciling (imperfectly) the needs of the recipient with the form of the data actually supplied. The supplier has often collected data for a different reason and is in any case constrained by practical factors associated with primary data collection, usually related to economics and technological capability.

It is reasonable to suppose that attempting to formalise the semantics of the supplier and the consumer will help to reduce the manual effort required by enabling at least some of the processes to be automated. Indeed there is currently a significant amount of effort being expended in the development of explicit and formal representations of domain semantics. The vehicle used for the representation is an ontology. Similarly, significant amounts of work are being conducted (most visibly by the OpenGIS Consortium (OGC)) into the syntactic description of geographic entities represented as digital features. Our own research interests, undoubtedly influenced by our role as a primary data suppler, are concerned with the relationships that exist between the ontologic concepts and the feature and in particular how the practicalities of data collection and supply combined with multiple world views disaggregate the feature and therefore complicate the concept/feature relationship. This paper concentrates on the manner in which features are represented and modelled, touching on the ontological aspects to provide a proper context.

2 Semantics and Ontologies

To enable sharing of data in a distributed environment with multiple world views requires adequate formal descriptions of those views and automated machine reasoning and manipulation of the data on the humans' behalf. The formal description and specification of different conceptualisations of the world is possible in ontologies which encode the concepts of a domain and the relationships between them. Through the modelling of an information community within which these are valid, a domain ontology expresses the elements which are to be represented. Indirectly, an ontology provides a formal description of the data required model concepts and does so independently of the internal or application-specific data-structures.

The encoding of logical relationships between concepts, through concept properties and axioms in ontologies, allows machine reasoning to deduce new knowledge about what is represented. Such relationships may include equivalence of class or property or whether properties are transitive or inverse. From these base constructs it is possible to build more complex relationships such as containment, part of, etc. This allows the possibility of selecting the concepts required for a given problem or domain to occur through automated reasoning and deduction carried out on the humans' behalf. This is particularly important in looking to transfer between static information communities. This may be by translation between ontologies (Rodriguez and Egenhofer 2003), or using different levels of related ontologies (Fonseca *et al*. 2002) or by integrating ontologies.

The use of ontologies for intelligent reasoning about the selection of appropriate data for a given task is also being researched within OGC catalogue services (Vögele, 2003) using spatial and temporal ontologies. The ability to describe the semantics of services that provide or manipulate data is also emerging in a number of languages such as DAML-S, (Einspanier et *al*., 2003). With automated inference against these descriptions the potential could rise for the intelligent chaining of data services.

The essential aspect is that it is increasingly possible to specify what we actually mean, not through simple syntax but in formally defined ontological concepts. The different approaches specify the meaning of what is required at different levels of granularity and for overlapping world views. This leads to the requirement to relate these concepts to the data captured about geographical features at the appropriate level.

3 Features

The basic unit of geographic information within most models of geographic information is the 'feature', where a feature is an abstraction of a real world phenomenon and a geographic feature is a feature associated with a position on earth (OGC, 2003, 3.1). Features can include representations of a wide range of phenomena that can be located in time and space such as buildings, towns and villages or a geometric network, geo-referenced image, pixel, or thematic overlay. This means that traditionally a feature encapsulates all that a given domain considers about a single geographic phenomenon in one entity.

Features can be seen at two levels *feature instances* and *feature types*. Feature instances are the individual discrete representations of geographic phenomena in the database with geographic and temporal dimensions. The instances may then be grouped into classes with common characteristics to form feature types. In OGC terms however features are not fixed in their class but have application-oriented views that are classed (OGC, 2003, 3.1.2) that is to say one view or application may class an instance as house while another may class it as a building. It is apparent then that the feature is not the atomic unit of geographic information as the phenomenon it represents encapsulates differing human concepts resulting in multiple types.

A closer examination of the behaviour of those involved in the collection, management, supply and exploitation of geographical information shows that different organisations utilise different collections of discrete sets of information about geographical entities. In this analysis the feature is a higher order representation of a shared understanding of a specific geographical representation and the actual atomic units of exchange are the discrete sets of information.

It therefore follows that in order to enable the exchange of feature based geographical information a model is required that allows the comparison of features not only as whole units but at a finer level that can consider the difference in what they encapsulate. The data model would need to describe the *partonomy* of the data in meaningful ontological concepts so that a more flexible approach can be adopted in managing the classification and taxonomy of the data.

4 Data model

Several authors have recognised the difficulty in reconciling differences in feature definition and have proposed, for varying reasons, modelling or defining geographical entities at a sub-feature level to allow the explicit management of varying feature definitions. Rodriguez and Egenhofer (2003) examined modelling geographic entities in ontologies in terms of different *distinguishing features* in order to enable correspondence between different entity classes to be evaluated. Usery *et al*. (2002) disaggregate the feature to enable multiple representations. Brodaric and Hastings (2002) have an implicit representation of features to allow interpretation when classifying geological information by taking into account the uncertainly in measurement and human judgement in the classification of rock formations which may be instances or evidence of existence. The 'component based' approach to feature modelling of Hart and Greenwood (2003) is a meta-level model where features are modelled as a set of discrete components of *identity* and *data* with an ontological framework for the description and logical combination of components. A data component represents a discrete package of data that corresponds to a single concept defined in the ontology. The feature is modelled as a collection of data components and a unique, unambiguous and persistent identity component as shown by Figure 1.

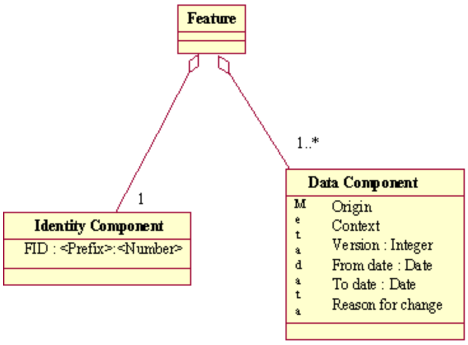

Figure 1. A simple UML diagram showing how the geographic feature is modelled as an aggregation of identity and data components.

This approach provides the potential to:

- Enable different data suppliers to gather different components representing different world views and assign them to the same feature;
- Manage this data at the most appropriate level – that of the data component, so change metadata, ownership (including versioning if required) and quality data is held at this level, not that of the feature.
- Make it possible for the task of data sharing to be carried out through a selection by the consumer (or end chain supplier) of the data components held by one or more data suppliers that correspond to the feature of interest.
- Enable the classification of the features based on the needs of the consumer since the classification can be derived from the supplied components as these are directly related to the ontological framework. This relaxes the requirement of having to define all feature types *a priori* and also enables new world views to be constructed by a data supplier or consumer
- The differences in domain definitions of features may also be examined as the difference in the set of associated components used.

5 The Logical Model

5.1 Overview

As the flow diagram in Figure 2 shows graphically, the logical model needs to be able to return typed feature instances corresponding to a set of ontological concepts that have been determined through human interaction and ontological reasoning by a client application. In order to enable the correct feature type to be returned, a clear set of relationships needs to exist between the concepts of the ontology, the data components in the database and the set of identity components which represent unique or agreed correlations between measured concepts, and so represent distinct features. This set of relationships will allow the selection of data components to create the required feature type definitions and corresponding database views to retrieve the required instances.

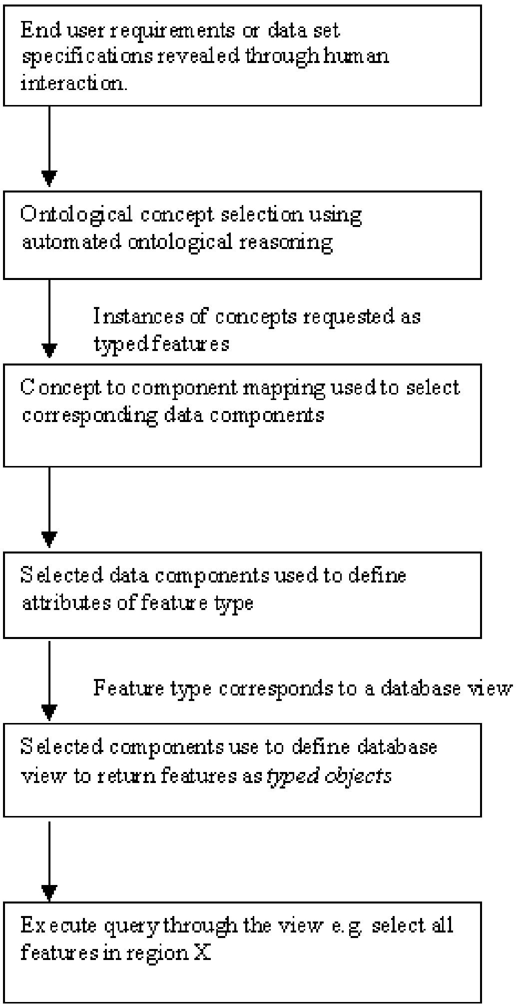

Figure 2 A flow diagram showing the sequence of steps required to progress from the selection of ontological concepts to retrieving typed instances of features.

5.2 Identity Components

These are modelled as typed objects consisting of the organisation prefix and the organisationally local unique number for the feature (which together comprise the Feature Identifier or FID). Where the feature has been correlated with a feature held by another organisation and established as representing the same real world object, the identity component should also hold metadata concerning the correlation including; the process used to correlate, the date when correlation occurred, a measure of certainty for the correlation and cross references to other identities the feature may have.[1]

The identity components map to a simple object table holding the machine-generated instances. References to the associated components are not required as the identifier are essentially non-human settable, they have no meaning as a point of query. To find the instances of the required component types would involve testing every reference held by every identity component for the type it represents. The reference is therefore held by the data components and is unidirectional.

[1] The complexity of the identity component is not discussed further here, nor is the form of the identifier nor rules for the allocation of identity. Whilst the latter two are not difficult to resolve technically they are often the subject of much political and commercial rivalry. We leave these areas to a later paper.

5.3 Data Components

Data components represent the basic units of information in the data model and correspond to formally defined ontological concepts. They are the digital measurements of an instance of the concept. Different types of component may represent the same concept such as extent at different scales. Components in the logical model are defined as objects and map to object tables, such as Table 1 below, where each row of the table corresponds to a single instance with each column representing an attribute of the component. The metadata is kept as part of the components in this table. Versioning also occurs at this level. Strictly the version may not be necessary as the information can be computed from the date ranges of successive component instances. To reduce redundancy in metadata information the values may be defined by reference to the set of metadata.

Table 1 Basic object table for a data component with a column corresponding to each attribute, and each row corresponding to an instance.

Origin	**Context**	**Version**	**From date**	**To date**	**Reason for change**	**ID_component**

Table 1 shows the object table for the basic data component, specializations of this would implement specializations of this table to include the additional information and relationships to other components that they support. There is, in theory, no need to assign a unique identifier to the component as uniqueness can be achieved through aggregating the FID, Origin, Context and Version. Although we have not done so it has been suggested that adding a component identifier will often be more convenient and practical.

5.4 Concept to Data Component Relationships

There is a one to many relationship between a concept and data the component type in that the component represents the specific set of measurable properties of one concept. There are at least three different ways in which this one to many relationship may come about.

1. The collection regime of a data supplier may require different lifecycles, for example height and building geometry may be captured using different capture techniques (SAR and ground survey) resulting in different lifecycles for the same building concept;
2. Different organisations may capture different information about the same concept;
3. The same set of properties may be captured at different precisions.

To implement this it is necessary to have a look up table, as shown in Table 2 below, so that given a set of required ontological concepts the corresponding data component types can be identified. The update of this mapping between the concept and the component is necessary in maintaining the relationship between the database and the ontology. The data type for storing the ontological concept will vary according to the format or language of the ontology being used. Retrieval of the required component types is then done using the SELECT statement to return the set.

Table 2 A simple look up table to hold the correspondence between ontological concepts and data components.

Concept	Component Type	Component Table
<concept 1>	component1_type	component1_table
<concept 2>	component2_type	component2_table

Potential[2] data component relationships and interdependencies such as topology and mereology will reflect those of the individual concepts they represent. Data components as instances of concepts implement these relations in full, independently of the features they are part of and get their identity from. The implementation and maintenance of the relationships defined between concepts in the ontology at the component level is not automatic nor is there currently an explicit mapping between concept relationships and the equivalent component relationships which instantiate them in the database. For example if the concept of a hospital defines "contains" relationships to the concepts of buildings there is no explicit mapping to the topological structures that return these. Where there is a need to build queries where concept relationships are constraints to that query then there is a need for these to be mapped to the database structures. This can be done by the introduction of associative components (special data components) that represent the relationships between concepts by referencing the appropriate data components or features that represent those concept instances. So the hospital feature has a data component containing merological references to the building features that are part of the hospital and perhaps also to features representing grounds and so on.

Relationships between features are thus determined by the relationships that exists between the components they possess and these in turn are defined by the concepts that are applicable to the real world object being described. Feature relationships are therefore ultimately determined by the ontological concepts they represent.

5.5 Data Component to Identity Component Relationships

A one to many relationship exists between the identity components and the data components. Each data component has a reference to the identity component to which it belongs. To retrieve all components (or a subset of them) requires a join on the object tables by the identifier component.

In some instances it may seem appropriate that a component may be part of more than one feature such as an address which can be seen as an attribute of different features, for instance it could be a component of a 'building' feature in one world view and a component of a different 'property' feature in another. The implication here is that two world views, and hence ontologies, share the same concept of address but associate this concept to different

[2] It should be remembered that an ontology defines the range of possible values and relationships that may exist. For example building will have a height and may have functions (roles) such as use as a factory associated with it. However, a database representation of this factory may have neither of these values or relationships expressed either because the appropriate fact is not known (the height of the building is unknown) or the value or relationships does not exist – the building is disused and has no function.

concepts in their respective ontologies, property in one cases and building in another. Such differences occur because of differences in geographical interest. A land registration body will be interested in the full extent of a property and hence will see the address as a reference for this whole. An insurer by contrast, will only be interested in the building on that property as this is what is insured. This would lead to a many to many relationship which would potentially be intractable in practical terms. It is our belief that many to many relationships should be resolved at the ontological level. In the address example, the relationship between one ontology and the other is that a building can form part of a property. Thus there is a relationship that can be mapped from one ontology to the other to enable address referencing to be resolved.

5.6 Creating Features Types

Given a set of required concepts the task is to be able to use the mapping between the concepts and the components to define and retrieve the typed features. This occurs by a client application selecting the set of ontological concepts required to be applied to the feature types through ontological reasoning or human interaction. The corresponding component types are retrieved from the mapping table (Table 2 above) and each component will correspond to an attribute of the feature type. Taking this set of retrieved component types, a stored function in the database (or client application) can be used to generate the required feature type definition. The attribute names correspond to the concepts and the data type corresponds to the component type as shown below using ORACLE's dialect of SQL:

```
CREATE TYPE feature_name AS OBJECT (
Prefix          string
ID              number
concept1        componet1
concept2        componet2
);
```

From the concept to component mapping, in Table 2 above, a view is then generated to specify the selection of data components in the same order as the attributes defined for the feature type (above) along with their storage location. This view specifies the constraint of a shared identity component between the data components.

```
CREATE VIEW feature_view OF feature_name WITH OBJECT IDENTIFIER (ID) AS
SELECT
        i.PREFIX, i.ID, VALUE(c1), VALUE(c2)
FROM identity_table i,  componet1_table c1, component2_table c2
WHERE VALUE(i) = c1.ID_component AND VALUE(i) = c2.ID_component
```

This potentially creates many joins across the required object tables with the resulting performance overhead. As identity is a specialist data type a custom method of comparison would be necessary (MAP or ORDER method in ORACLE).

5.7 Retrieving Instances of Feature Types

The selection of feature instances can now be done through the view, using the SELECT command, which specifies and manages the required shared identity reference between components. Any required geographic constraints are then applied to queries through this view.

5.8 Summary of Logical Model

This logical model allows the implementation of a partonomic description of features in ontological terms. This is achieved through the use of a mapping between ontological concepts and the components of features. This mapping can then used to dynamically generate the required feature types and instances through the use of database type and view mechanisms. This logical model is designed to allow the continued exploration of the relationship between ontologies and feature based geographic data rather than being a definitive commercial implementation.

6 Conclusions and Further Work

This paper has looked at the role the feature plays in the sharing of geographic information and the fact that a feature encapsulates different human concepts depending on its application. It has shown how a partonomic description of features using data components can be linked to ontological concepts to allow the flexible description and selection of features. The combination of the partonomic model and ontological concepts provide the potential for a more dynamic approach to the classification of features and feature types.

Further work is required to show how concept relationships can be mapped from the ontology into relationships between data elements such as topological constraints. Further work in defining user scenarios will enable the identification of the types of queries that depend on explicit relations and help determine the logical structures required. Tools are also required to keep the ontology and ontological mapping to the data components in synchronisation and to allow for easy update and authoring.

7 References

BRODARIC, B., and HASTINGS, J., 2002, An object model for geologic map information. In: *Proceedings Spatial Data Handling 2002*, July 9-12, Ottawa, Canada.

BURROUGH, P., and MASSER, I., 1998, International aspects of spatial data exchange. In P. Burrough and I. Masser (Eds.) *GISDATA 5 European Geographic Information Infrastructures opportunities and pitfalls.* pp. 1-14. Taylor and Francis, London, 1998.

EINSPANIER, U., LUTZ, M., SIMONIS, I., SENKLER, K. and SLIWINSKI, A., 2003, Toward a process model for GI service composition. In L. Bernard, S. Sliwinsk und K. Senkler (Hrsg.) *Geodaten- und Geodienste-Infrastrukturen – von der Forschug zur praktischen Anwerdung, Beiträge zu den Münsteraner GI-Tagen.* 26-27 June 2003, Münster, Germany. 2003.

FONSECA, F. T., EGENHOFER, M. J., DAVIES, D., and GILBERTO, C., 2002, Semantic granularity in ontology-driven geographic information systems. *Annals of Mathematics and Artificial Intelligence* Vol. 36 No 1-2 pp. 121-151.

HART, G., and GREENWOOD, J., 2003 A component based approach to geo-ontologies and geodata modelling to enable data sharing, In: *Proceedings of the 6th AGILE,* April 24-26, 2003 Lyon, France,

OPEN GIS CONSORTIUM , 2003, *OpenGIS Reference Model,* Version 0.1.2, Open GIS Consortium Inc. Wayland, MA, USA.

RODRÍGUEZ, A. M. and EGENHOFER, M. J., 2003, Determining semantic similarity amoung entity classes from different ontologies. *IEEE Transactions on Knowledge and Data Engineering,* Vol. 15, No. 3, pp442-456.

USERY, E. L., TIMSON, G., and COLETTI, A., 2002, Multidimensional Geographic Feature System. In G. Zavala (Ed) *GIScience 2002 Abstracts. GIScience 2002 The Second International Conference on Geographic Information Science.* Boulder, Colorado, USA pp 279-282. University of California Regents

VÖGELE, T., SPITTEL, R., VISSER, U., and HUBNER, S., 2003, GeoShare – Building a Trans-national Geodata Infrastructure for the North Sea Reagion In L. Bernard, S. Sliwinsk und K. Senkler (Hrsg.) *Geodaten- und Geodienste-Infrastrukturen – von der Forschug zur praktischen Anwerdung, Beiträge zu den Münsteraner GI-Tagen.* 26-27 June 2003, Münster, Germany. 2003.