Proceedings of the 6th AGILE
April 24th-26th, 2003 – Lyon, France

# A COMPONENT BASED APPROACH TO GEO-ONTOLOGIES AND GEODATA MODELLING TO ENABLE DATA SHARING

**Glen Hart and Joseph Greenwood**

Research & Innovation, Ordnance Survey of Great Britain,

Southampton, Hampshire, SO16 4GU, England

[ghart, jgreenwood]@ordsvy.gov.uk

www.ordnancesurvey.co.uk

## 1. INTRODUCTION

There is a growing requirement for a more integrated approach to exchanging data between organisations, whether this is to enable web and location based services or provide seamless inter-business and inter-government processes. It is apparent that technology is not the major restriction. Excluding commercial or political impedance, it is the reconciliation of differing world views that presents a major barrier to data interchange. Burrough and Masser [1] note that it is "the fact that related sciences and technical methods do not necessarily describe or recognise the same spatial phenomena in the same way". These disparities result from differences in semantic, functional, temporal and locational understanding that are expressed in the separate knowledge domains (world views). Often the features of interest are close enough in definition to warrant sharing information. And, if not, the actual differences in description are of interest in understanding different stakeholder views and which through combination may provide others with a more complete picture than would otherwise be possible.

Most success in sharing data has been in well-defined areas and applications, such as those described in Langaas [2] where although world views differ they are complementary in that they either enrich each other for the same geographic area or the aggregation provide a larger geographic coverage. They thus enable interoperability with minimal error or for a compromise standard to be agreed (as with the British National Land Information System, NLIS, www.nlis.org.uk). However, the current form of databases or frameworks for sharing data, limits their ability to meet changing world views and business processes and limits the ability for such systems to increase the number of contributors. In contrast, this paper describes an approach that reveals and manages the multiplicity inherent in differing world views to maximise the data interchange and use between those views.

The use of data specification standards in isolation is recognised as insufficient (Fonseca et *al* [3]), they can act like a barrier between different communities in which they dominate, do not necessarily provide a rich semantic or functional description of the data they contain and are restricted in the range of concepts they can convey. By their definition they are fixed in their possible interpretations of the real world. Consequently, we are not seeking to develop and apply a single dominant or canonical view of the world but an acceptance and management of them.

The paper outlines the progress we have made in exploring, at conceptual level, a framework for the development of "open"[1] geo-ontologies and geodata models in an integrated manner.

## 2. APPROACH

The approach outlined contributes to the architecture and standards goals of INSPIRE for the publication of metadata and data identified as the highest priority for a spatial information framework (INSPIRE 2002 2.2.1.7) [4].

Our method is to develop an ontology that provides a more open framework for describing a world view than many existing approaches. Concepts[2] defined within the ontology are used to determine discrete packages of data of a general geodata model. We call these packages '*components'*. The geo-ontology is then used to constrain the application of this geodata model. We will show how such an approach can be exploited by multiple data suppliers and users each of which may have differing world views. This is achieved through the use of discrete components of *identity* and *data* and an ontological framework for their description and logical combination. This formalises the use of unique identifiers (identified as a research priority in INSPIRE 2002 4.7 [4]), data and ontology.

In essence, whereas initiatives such as OpenGIS and ISO/TC211 place emphasis on the *feature* as a whole and where the feature is a digital representation of a "real world object or phenomena" (RWO), our emphasis is has been to focus at a sub-feature level (component) as we believe that this enables much greater flexibility in expressing the richness of the world around us. In correspondence Broderick has stated "OpenGIS [SIC] and ISO/ [SIC] TC211 are mainly feature-centric, where feature-types (i.e. low-level concepts) are defined beforehand in a data model and the link between features and feature-types is instantiation. Firstly, this setup favors static feature-types, secondly, its difficult to assign multiple world views (i.e. feature-types) to a feature. This is not optimum for scientific work where feature-types are evolved as part of the scientific process, implying more complex relationships between evidence data (features), concepts (feature-types) and instances (features). We discuss these relations in our GISci'02 paper {[5]}, though not in terms of OGC and TC211". We support this view and although here he speaks in terms of science and the evolution of schemas (feature types) we believe that the approach (OpenGIS and ISO/TC211) is also not optimum for most situations where multiple world views are coalesced and where RWOs change in terms of definition. This does not mean that we are unsupportive of OpenGIS, ISO/TC211 or the technologies such as OO upon which they are founded, more that we perceive limitations where multiple world views are required with these otherwise successful initiatives and technologies. We therefore seek to extend the principles upon which they are founded rather than replace them.

In truth, a feature is not strictly a digital representation of a RWO but a collection of data (fact and inference) that is believed to be true and which is determined by the world view of the collector and the collection regimes operated by that collector. One aspect of our approach is to more explicitly recognise this and also to recognise that a RWO may be represented by more than one discrete packet of information.

[1] By this we mean that they are not restricted to a single world view.
[2] Our working definition of a concept is "A notion formed in the mind representing an abstraction of an aspect of a real world feature. For example a newsagents shop might be described using concepts such as building, shop, postal address, etc. "

We also believe that the principles of our approach can be exploited to support concept evolution in a similar fashion to those explored by Broderick and Gahegan [5]. We do not investigate this further in this paper although it is our intention to do so in the near future as a necessary mechanism for managing changing world views.

Using ontologies as a foundation to data sharing has been previously suggested, but authors have often either left the relationship to the data model undefined, (Kuhn [6]) or have tied it to one physical implementation method such as object-orientation, Fonseca et *al*. [7] for example.

Our approach has been to develop a meta-geodata model where the precise shape and form of a particular geodata model is described through an ontology. Whereas much current research in ontologies has been directed at the semantic aspect we have concentrated on exploring the ability of an ontology to explicitly describe concepts and their relationships to both make visible to others a particular world view and, inwardly, to enable these definitions and relationships to dictate the building blocks of a geodata model.

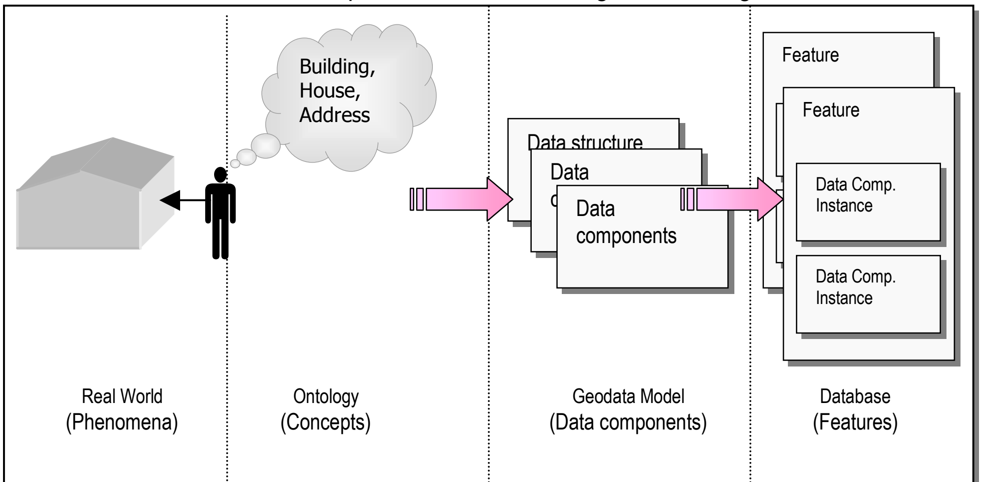


**Fig. 1** Relationships, real world to Database

## 3. GEO-ONTOLOGIES AND GEODATA MODELS

We model the world as a set of concepts that RWO possess (or that we infer as applying to RWOs). These concepts in turn relate to the discrete data components within the geodata model and are realised as features (collections of instantiated data components associated with an identity component) in a database. This is shown in Fig 1.

### 3.1 Geo-ontologies

We identify and describe these concepts in a geo-ontology in the following terms:

- Natural language description;
- Physically measurable (or estimable) properties including geometry;
- Relationships to other concepts including:
  - Mereology and Dependencies of existence (e.g. apartment and building);
  - Topology;
  - Taxonomic relationships (specialisation, generalisation).

This ontological description allows us to identify the relevant properties of the concept that are of interest and the relationships that we believe the concept has.

Examples of concepts include geographical form such as premises, building, river, road; function such as shop, office, factory, pasture; and properties such as address, name, and extent.

We have not attempted an axiomatic description of concepts although acknowledge the importance of this aspect (and also its complexity).

Concepts in the ontology are used to describe RWOs. So the RWO: "The Mended Drum Public House, Ankh Morpork, Discworld"[3] could be described by concepts (if they are defined) such as Premises, Pub, and Home (if it is where the Publican resides). The Premises concept might have relationships and dependencies to the RWOs from which it is comprised (the buildings, a beer garden) and perhaps topologic relationships to indicate it has road access. It will also "possess" as measurable properties the concepts of address and geographic extent. Likewise, the concept Pub may define properties such as opening hours, range of wines, beers and spirits and taxonomic relationships such as a specialisation of "Eating and Drinking Establishments" (or from a different world perspective "Likely Trouble Spot").

In this manner detailed descriptions can be built up of individual RWOs by combining different concepts through aggregation. The approach is therefore an atypical ontological approach in that it is not hierarchical. For example Kuhn [6] demonstrates a similar combinatorial approach but uses inheritance as the mechanism to achieve the "blending". Our approach avoids the problems inherent in a hierarchical approach such as conflicts of multiple inheritance and the need to guess *a priori* all allowable classifications that will be encountered, delaying the point of decision from ontology development to feature instantiation and modification. Kuhn also introduces the notion of "*image schemata*", "basic conceptual structures rooted in the human perceptual system, to capture ideas like containment, support…". This notion is absent from our work but could well be a fruitful path to follow for grounding our concepts in a semantically meaningful way.

### 3.2 Geodata Model

The data model describes features (digital representations of RWOs) in terms of identity and data, each held in discrete component form. A feature may be viewed as a digital identifier component associated with a number of data components (Fig 2). This idea is similar to concepts expressed by Usery *et al*. [8]. although here applied in a wider context than to multidimensional representations.

The geodata model itself only describes the form of the identity component, the specific structural forms of the data components (as

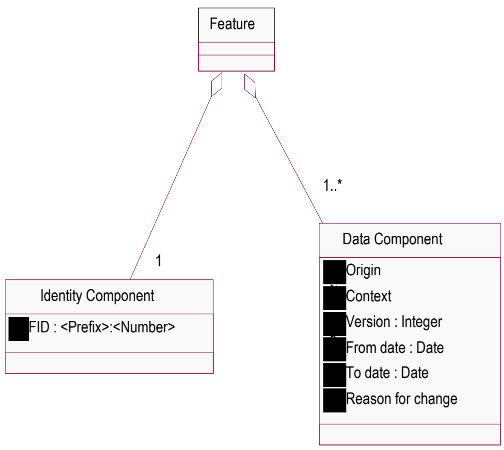


**Fig. 2** - Feature Metamodel

[3] Acknowledgement to Terry Pratchett and his Discworld novels for this example. The Mended Drum, Discworld® and Ankh Morpork © Terry and Lyn Pratchett.

specialisations of the general data component) and the relationships that exist between component types.

The identity component holds a machine-generated Feature Identifier (FID), which contains no intelligence about the feature and acts to represent the notion of existence of the feature. It is unique and exists in perpetuity.

The general form of the data component defines the metadata necessary to manage the data associated with the component including originator, record context and time stamping[4]. Other metadata that may be specific to a particular concept related to a feature would be held within the specialisation of the data component. For example, a data component representing the concept of 'spatial extent of a premises' might contain metadata identifying how the extent had been surveyed and the degree of accuracy achieved.

### 3.3 Concepts revisited

From the above it should be apparent that the concept of "concept" is not itself an atomic unit. Just as we argue that identity and data components are the atomic units of the feature, we also believe that a "conceptual component" is the atomic unit at the conceptual level and indeed that there is even a case for a corresponding Concept Identifier. The concept component would represent the interpretation of a shared concept from a specific world view, i.e. the concept of a house as defined by a taxation body or from the view point of a mapping agency. In the rest of this paper we gloss over this additional complexity ashamedly because whilst recognising its importance we have not yet develop the idea further although we are again pointed to the work of Broderick [5].

Whilst only touching on this aspect in this paper we believe this structural symmetry between concept, concept component, data component and feature is important in ultimately being able to make sense of multiple world views.

### 3.4 Geodata Model Relationships

The principal relationships between the geo-ontology and the geodata model are:

- Concepts determine the data components that may exist within the data model
- Relationships between concepts expressed within the geo-ontology determine the relationships that may exist between data components.

The relationship between the data model and the database is a little simpler as a feature (an instance within a database) describes its associated RWO through any combination of data components. The full range of applicable components are determined by the concepts associated with that RWO and the actual components present are a subset of the possible range determined by what is actually known about the RWO.

In returning to the relationship between concept and component it is our belief that such a relationship is one to many although we still need to firmly establish this. The factors determining the existence of a data component include:

- Data that would be associated with the parent concept;
- The theoretical lifecycle of the concept modified by the practical lifecycle of the information collection regime(s). This nicely emphasises the difference between

[4] In a similar manner to the approach of Usery *et al* [8] it is the data components that represent the units of change and are version managed. This enables histories to be constructed not of the feature but of the information known about concepts that are associated with the feature.

the RWO and the information that maybe known about the RWO at any time (which is what is modelled).

- Differing world views that recognise the same concept but whose information needs are different.
- Differing requirements of end-use, for example the extent of a premises could be represented by geometries at different scales, each represented by a different data component.

This one to many relationship between a concept and data components reduces the ambiguity in determining the contents of an instance of a data component as it is the discrete set of data for that *one* concept as opposed to many concepts, ensuring its validity as an atomic piece of data. The need for multiple concepts, or multiple inheritance in an object-oriented paradigm, to describe a single RWO is handled by the relationships defined between concepts in the geo-ontology used to determine the aggregation of data components to make a feature. Multiplicity is explicitly handled once at the definition level rather than implicitly at the data instance level

Data component relationships and interdependencies such as topology and mereology will reflect those of the individual concepts they represent. Data components as (part) instances of concepts implement these relations in full, independently of the features they are part of and get their identity from. Relationships between feature instances 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 RWO being described. Feature relationships are therefore ultimately determined by the concepts they do or do not share.

The geo-ontology is thus able to constrain the very open nature of the geodata model[5] through constraining the allowable relationships that may exists between concepts and in their physical manifestation as data component instances.

### 3.5 Feature Classification

The classification of the feature can be derived interpreting the concepts that are known to apply to a feature. So the pub in Ankh Morpork is classified as a Building, a Pub, a Home, and so on. This approach is clearly more flexible than the existing feature typing and has three distinct advantages:

- Different combinations of concepts enable a very rich set of realities to be described;
- Where concepts are found lacking (a ‘new’ concept is discovered) a new concept can be introduced and new data components developed. This is a additive process and would not alter existing definitions or structures;
- It is avoids to a far greater extent the issues surrounding conflict that are associated with using hierarchical typing systems..

Hence, it is no longer necessary to construct hierarchical compromises with the inherent inflexibility that they bring. Static feature typing is no longer a restriction and a feature maybe completely reclassified to reflect real lifecycle changes to the RWO.

[5] This openness being the models strength and, if uncontrolled, its greatest weakness.

## 4. PLURALISTIC WORLD VIEWS

The approach described above, largely from the perspective of a single world view, allows the sharing of data through the use of common parts of an ontology that is sufficiently rich to enable correlation of concepts.

With this approach there is no reason why the conceptual components need be extracted from only one world view, nor why different data components applicable to the same concept may not be provided.

This enables new concepts to be introduced and definitions of existing concepts to be extended. In this way a feature representing a RWO may be described using differing world views, each of which adds to the richness of description.

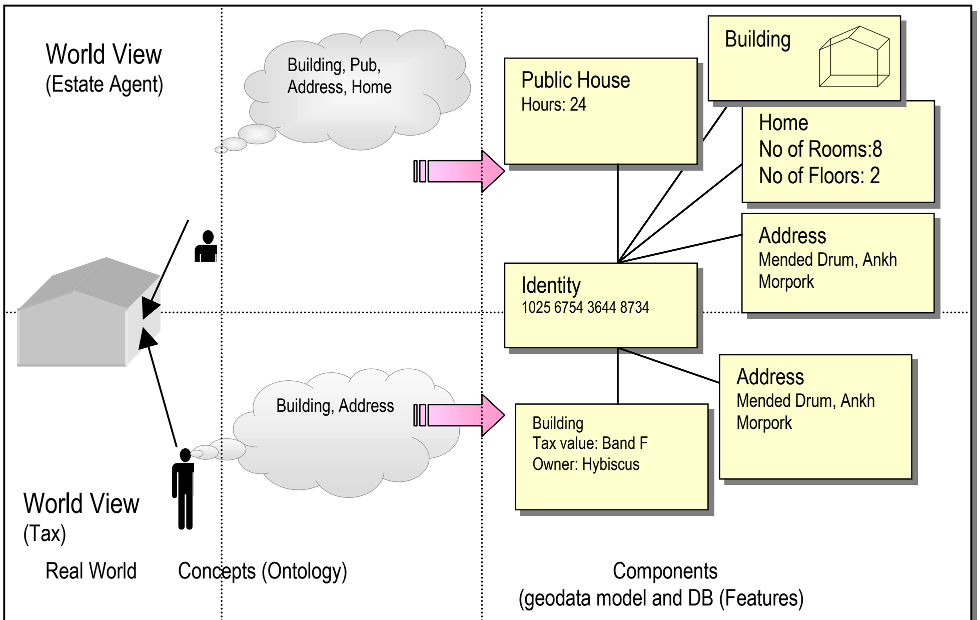


**Fig. 3** Two World Views of the same RWO

Fig 3 gives a simple example where two different world views of fundamentally the same RWO can be correlated[6]. One view, that of an estate agent, has a particular conceptual view of the pub in Ankh Morpork. In this view the concepts of building, pub, address and home are seen to be important. In terms of the information held, data components for each concept are defined and populated (opening hours against *Pub*, Address details for *Address*, building model for *Building* etc.). The view of a tax inspector may be somewhat simpler – essentially the taxable value of *Building* and an address. In this simple example (we don't for example show or discuss relationships) the concept *Address* is shared exactly by the two world views although the data may be held independently in separate databases. The Address concept because it is a shared concept can be used to correlate the different views of the same feature. The world views extend each others

[6] For drawing convenience the Geodata model and Database views, shown separately in Fig 1, have been combined in this diagram.

concept of *Building;* and *Pub* and *Home* are concepts only of relevance to the first view. Through correlation each organisation achieves a shared view of a feature.

Feature relationships are therefore the result of concepts they do or do not share. For such a method to work there has to be:

- sufficient understanding of the concepts that are defined in an ontology to enable another party to add new concepts and extend existing concepts;
- The ability to enable a new party to extend the geodata model by adding new components;
- Sufficient information within databases to enable features to be correlated (or on receipt of more accurate information, to decorrelate). As the problem space is geographic it is suggested that a sufficient amount of information (as represented in discrete data components) would comprise some sense of classification, location and time (what, where, when);
- An agreed infrastructure to enable correlation and decorrelation of disparate data including a robust means to allocate unique identifiers.

Data sharing thus becomes possible through this mechanism. For this to happen an exchange must occur of:

- the data;
- a description of the data through meta-data and the concepts of the components. This metadata is held with the geo-ontology for conceptual descriptions and within the geodata model where descriptions related to the logical structure (components) are held);
- a context for that data. This is to say that the specific world views that are of interest must be declared, as an expression of the applicable concepts and the information the recipient is interested in (a subset expressed as required data components[7]).

Where representations of a feature have been correlated, the FID acts as a *lingua franca* enabling data components to be exchanged based on shared identity. Where a third party wishes to access data and the FID is unknown to that party, it will be necessary for them to supply search criteria and a context to enable a consistent set of data to be retrieved and presented.

There are undoubtedly dangers with this approach, even excluding the potential for miscorrelation. A plurality of world views brings with it at least an equal number of contexts and it is clear that it is possible that conflicting or differing data supplied by different parties may present ambiguities unless context is applied when interpreting the feature. Although tied to the object-oriented approach the idea of *role* as explored by Fonseca *et al.* [9] may help to resolve this.

Other problems arise from misunderstanding existing contexts and thus extending a current conceptual definition rather than introducing new ones.

However, where organisations wish to share information all these problems exist irrespective of any formalised method. What we are aiming to achieve is to make such potential conflicts explicit, enabling techniques for their management to be developed over time.

[7] We are regretfully vague here. A personal belief is that such context can be achieve by the definition of the concepts which are of interest to the particular world view, but also that a finer level of granularity is required (i.e. just as we model features in component terms we also need to model concepts in terms of conceptual components.) More work, more work!

## 5. CONCLUSION

This paper has outlined a method of expressing geo-ontologies and geodata models in a consistent and integrated manner. It has outlined a component based approach where database features are expressed using aggregations of data components described in the geodata model and associated through the use of a shared FID. These data components in their turn are defined by the geo-ontology which defines or describes supported concepts of individual real world views.

This paper has illustrated that the combination of an ontological framework and a component based model of identity and data has potential to accommodate multiple world views. It has elaborated on the relationships that need to be present in the geodata model, the ontology and between the ontology and geodata model (through the data components) to enable correlation between multiple world views**.**

The approach outlined shows promise, in that it has the potential to free ontologies from the constraints of hierarchies and extends this freedom to the related geodata models and ultimately the underlying databases. It also has associated dangers, particularly from the uncontrolled development of very open solutions with the potential for both over-complexity and possible semantic mis-match.

The development of working models and an understanding of how to manage the potential pitfalls are our immediate research goals.

## 6. REFERENCES

[1] 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.

[2] Langaas, S., 1998, Transboundary European GIS databases: review of the Baltic region experiences. In P. Burrough and I. Masser (Eds.) *GISDATA 5 European Geographic Information Infrastructures opportunities and pitfalls.* pp 31-43. Taylor and Francis, London, 1998.

[3] Fonseca, F. T., Egenhofer, M. J., Agouris, P. and Câmara, G., 2002, Using ontologies for integrated geographic information systems. *Transactions in GIS* Vol. 6, No. 3 pp 231-257, 2002.

[4] INSPIRE, 2002, *INSPIRE Architecture and Standards Position Paper*. JRC-Institute for Environment and Sustainability, Ispra.

[5] Brodaric, B. and Gahegan, M., 2002, Distinguishing Instances and Evidence of Geographical Concepts. In M. J. Egenhofer and D. M. Mark (Eds) GIScience 2002, LNCS 2478 pp 22-37, 2002. Springer-Verlag, Berlin, 2002.

[6] Kuhn, W, Modelling the Semantics of Geographic Categories through Conceptual Integration, In M. J. Egenhofer and D. M. Mark (Eds) GIScience 2002, LNCS 2478 pp 108-118, 2002. Springer-Verlag, Berlin, 2002.

[7] Fonseca, F. T., Egenhofer, M. J., Davis, C. A. and Borges, K. A. V., 2000, Ontologies and knowledge sharing in urban GIS. *Computers, Environment and Urban Systems* Vol. 24 No. 3, pp 251-272, 2000.

[8] 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.

[9] 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.