Modelling Geographic Phenomena at Multiple Levels of Detail

Omair Chaudhry & William Mackaness

Institute of Geography
School of GeoSciences,
University of Edinburgh
Institute of Geography
Drummond St
Edinburgh EH8 9XP
O.Chaudhry@sms.ed.ac.uk
William.mackaness@ed.ac.uk

Keywords: Model Generalisation, MRDB, Spatial Analysis, Data Modelling, Semantics

Abstract

Many tasks require us to view the world at various levels of detail across a range of scales. Mapping agencies are keen to avoid the redundancy and cost associated with maintaining multiple independent databases. Considerable interest remains in capturing once geographical information at the fine scale, and from this, automatically deriving information at various levels of detail and scale. Prior to the cartographic portrayal of that information, model generalization is required to derive the higher order phenomenon associated with smaller scales. We argue that successful derivation of these phenomena requires us to model and make explicit some of the taxonomic and partonomic relationships that exist between geographic phenomenon. The paper reports on attempts to create a multiple representation database supporting the automatic derivation of objects found at 1:250,000 from a fine scale database at Ordnance Survey (OS) MasterMap (1:1250/10,000). Results from implementation are presented. It is argued that the value of such work extends beyond the visual and has important implications for spatial query and exploratory data analysis.

1.0 Introduction

Spatial data portrayed at multiple scales in map form has existed for thousands of years (Turnbull 1989). Cartographers have long understood the link between scale and task, (Monmonier 1984) arguing that it is a travesty not to supply mapping at multiple scales. There is a link between scale, the phenomena being represented and task. It is not that we see *less* information at coarser scale but that we see *different* information. For instance for pedestrian navigation within a city we require spatial data at large scale (showing detail) since it contains information at the street level; for navigation between or across cities a coarser view is required for the purposes of planning and to gain a better sense of overall distance and direction. Thus there is a need to represent spatial data at different levels of details to discern fundamentally different processes and patterns both in qualitative form (maps) and for quantitative analysis including Exploratory Data Analysis (EDA).

To satisfy this demand, National Mapping Agencies (NMAs) are creating spatial databases that support maintenance at the fine scale, and from which multiple products can be produced (varying in theme and scale/ level of detail). These databases are called Multiple Representation Databases (MRDB) since they store and represent various geographic phenomena at varying level of details (João 1998). Without intelligent links between the various phenomena contained within MRDB, maintenance costs are high, requiring updates to be applied across the scales, rather than at the fine scale, with generalised revisions automatically rippling through to lower levels of detail (Kilpelainen and Sajakoski 1995). There is much debate about ideas of 'scale' from a database perspective (Goodchild and Proctor 1997) and arguments about 'scaleless' databases, but the reality is that scale is inextricably linked to the phenomena being modelled (Levin 1992) and the data model in which the phenomena are represented has an implicit scale associated with it. Indeed the term 'scale-less' is meaningless – both from a modelling and a visualisation perspective. Populating coarser levels within the MRDB is achieved via model generalisation – the process of abstraction and transformation of spatial information from the fine scale to a spectrum of higher order objects (Kilpelainen 1997) which then form a basis from which they can be visualised via cartographic generalisation techniques (Kilpelainen 1997; Weibel and Dutton 1999).

The aim of model generalisation is to transform objects in the source database into higher order objects for the target database at higher level of abstraction, whilst preserving their salient or characteristic qualities, thus revealing a different set of interdependencies, topological qualities and patterns among higher order objects. This differs from cartographic generalisation which aims to improve the visual effectiveness and readability of a map (Brassel and Weibel 1988). Model generalisation can be considered as a pre-process to cartographic generalisation (Kilpelainen 1997) – the two are inextricably entwined (Figure1). Model generalisation is an essential prerequisite for large changes in scale. This paper focuses on derivation of small scale databases (notional scale of 1:250,000) directly from a large scale database (1:1250 scale) via model generalisation. We argue here that the degree of automation depends on the sophistication of the underlying model.

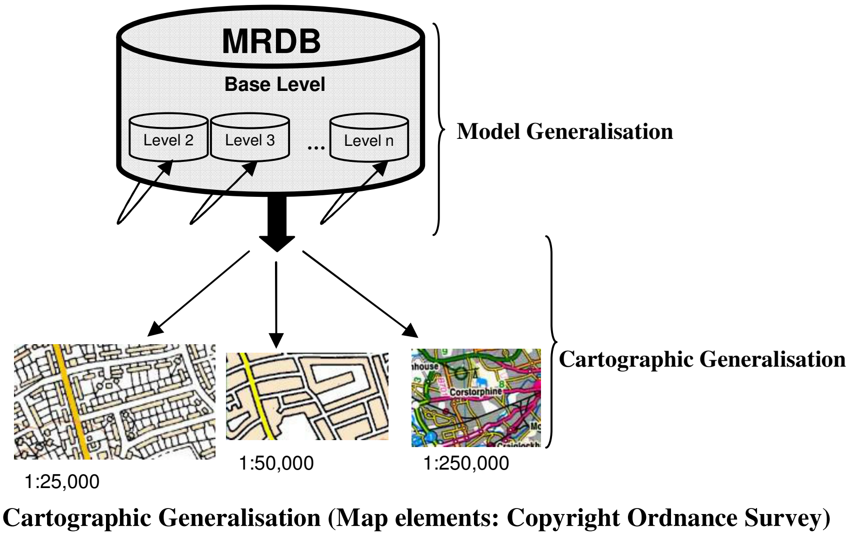

Figure 1: Model vs Cartographic Generalisation (Map elements: Copyright Ordnance Survey)

1.1 Problem Statement

Derivation of a small scale database is not straight forward and is more than just a process of *subselecting* the data. It involves creating fundamentally new geographical phenomena from given geographical entities. The transformation of the database involves creation of higher order objects such as cities, forest regions, and mountain ranges from lower order objects in the source database (such as buildings, trees and groups of hills). In this transformation process it is important to model phenomena in a meaningful way (Ormsby and Mackaness 1999), rather than to consider the object in terms of its geometric primitives (points, lines and polygons). Ideas of semantic modelling (Foerstner and Pluemer 1997), and the ability to characterise the saliency of objects at different levels of detail is critical to the interpretive process. That when the map reader sees a dot with the word 'London' next to it, they understand in an instant, what that dot represents, together with all the processes and phenomena that are contained within it. That 'dot' should be stored in a way that supports a whole set of meaningful queries and analysis techniques, quite separate from another dot, that say, represents the centre of a hurricane. It is also vital to understand the relationship between the higher order concepts and those objects at the fine scale; how objects contain other objects, or connect, or isolate objects. Modelling the changing topology is not something that has been sufficiently researched – but is critical to finding solutions to problems such as the one posed by Minsky – that 'You cannot tell you are on an island by looking at the pebbles on a beach'. Is it possible to develop automated generalisation solutions that can determine whether you are on an Island when all you have are pebbles?

1.2 Aim and Objectives

This research aims to derive a small scale (1:250,000) topographic database from OS MasterMap database via the process of model generalisation. We argue here that such a database is an essential precursor to more general forms of spatial query, and is an essential prerequisite to cartographic generalisation. The main objectives of the research are:

- To develop a model of the changing behavior of phenomenon over large changes in scale;
- To develop a data model that transforms one schema to another based on semantic modelling;
- To demonstrate through spatial query and visualisation – the form and extent of the newly derived objects;
- To evaluate the output beyond visual comparison using spatial analysis techniques.

Geographical modelling lies at the heart of this research. Success will enable us to query and analyse geographic data in a more intuitive way. The research will help us to appreciate the importance of meaningful concepts/phenomena in a spatial database rather than as a set of attributed points, lines or polygons and will also help us to obtain appropriate and intelligent answers to our questions instead of topologically correct but abstract answers.

The paper is organized as follows: Section 2 discusses the importance of semantic relations in database transformation; Section 3 gives overview of different phases of implementation; Section 4 presents the case study and discuses the utility of the results obtained. The paper concludes with thoughts on further research.

2.0 Semantics Modelling

In order to create new phenomena from source concepts it is critical to understand their semantics. This is because without looking into the meaning and relations of each object in the database we are just looking at a set of points, polygons and lines. Semantics of spatial objects is generally embedded in the context of the application (Molenaar 2004). To understand the semantics we need to understand the spatial context which is usually implicit in the data model. Most geographic data models explicitly represent spatial objects with their geometry and thematic properties (Mustiere and Moulin 2002), but to understand the semantics we also need to understand the topological and proximity relationships between objects (Worboys 1996; Mark 1999). The encoding of their relationships and properties will allow the deduction of new concepts required in the resultant database. There are two types of relations that we are particularly interested in: *taxonomic* and *partonomic* relationships – both are hierarchical in form and in both of these - relations, concepts and entities are represented in terms of class and objects. An object represents individual discrete geographic phenomena in a spatial database. Each object has a set of properties such as name, type, width, length. Objects that have a common set of attributes are grouped together to create a class. A class defines the attribute structure for each of its instances (objects). In other words all objects that belong to the same class will have the same attribute structure.

2.1 Taxonomic Relations

Objects having the same attribute structure are grouped together into a class and each object becomes an instance of the class it belongs to. This process is termed as classification. Classification is based on the thematic description of the objects. Thematic description of each object contains information about the state of the object or the role of the object in the database (Molenaar 1998). In other words this thematic description helps to understand the meaning of each object in terms of their attributes. Classification of these objects into classes helps to distinguish between different set of objects. In the context of topographic mapping, classification enables us to distinguish between objects based on their attribution. For example Figure 2 shows the different attributes of a house and factory class.

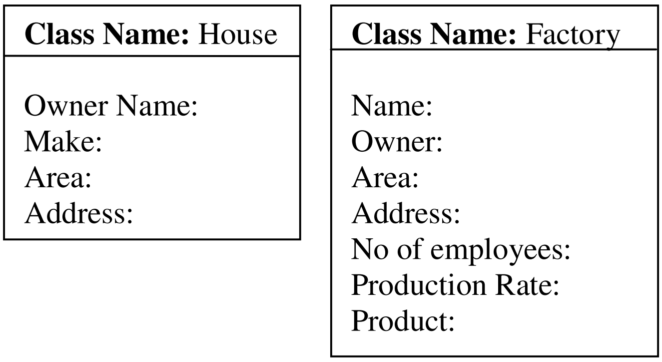

Figure 2: Classification

Different classes often have different attribute structures; though some of the attributes may be shared (we can create a *super class* whose attribute structure is made up of these shared attributes). The relationship between a class and its superclass is called *taxonomic or taxon* (Smaalen 2003). Classes and super classes forms a hierarchical structure called as *classification* *hierarchy* (Smith and Smith 1977; Thompson 1989; Molenaar and Richardson 1994; Peng 1997) or *taxonomy* (Smaalen 2003). An example from geological mapping is given in Figure 3, itself used as a basis for amalgamating rock units in categorical generalisation (Downs and Mackaness 2002).

A basis for classification and amalgamation:

Rock Unit Name (Lexicondes)	***Rock Type***	Parent Geology
Belsay Dean Limestone	Limestone	STAINMORE GROUP
Corbridge Limestone		
Dalton Limestone		
Great Limestone		
Little Limestone		
Stainmore Group		
Stainmore Group	Sandstone	
Rothley Grits		
Shaftoe Grits		
Eelwell Limestone	Limestone	Liddesdale Group
Lower Bath-House Wood Limestone		
Redhouse Burn Lower Limestone		
Redhouse Burn Middle Limestone		
Shotto Wood Limestone		
Upper Bath-House Wood Limestone		
Redhouse Burn Upper Limestone		
Liddesdale Group		
Liddesdale Group	Sandstone	

Figure 3: A classification schema used as a basis for categorical generalisation (Downs and Mackaness 2002).

Figure 4 shows parts of a transportation classification. One can readily envisage different map granularities associated with each 'layer' in the classification. Classes at different levels in the hierarchy correspond to data with various characteristics. If a database contained data at the finest scale, it should be possible, through the process of model generalisation, to populate the higher layers in this classification.

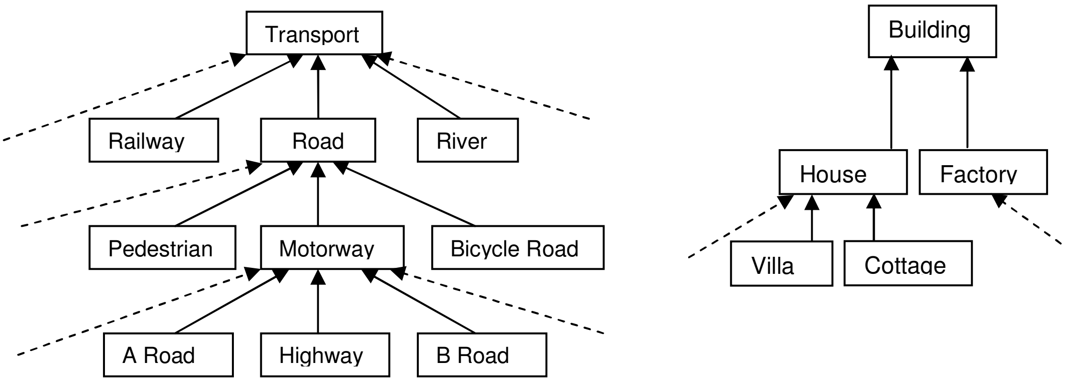

Figure 4: An example of a Classification Hierarchy or Taxonomy

2.2 Partonomic Relations

Based solely on taxonomic classifications we can reclassify phenomena belonging to the same classification (all types of woods, or all types of transport). But often we want to combine phenomenon from different classifications. We may want to combine phenomena based on their shared function. For instance a city might be made up of a density of roads, churches, industrial quarters, stations, and political institutions – it is what *defines* 'citiness' (from a prototypical and functional point of view). All these objects belong to different classification hierarchies, but when in physical proximity and density, it is valid to aggregate them and create a new object of class 'city'. A particular set of objects are 'part of' a particular instance of a city. It is interesting to note that an objects relationship with respect to other objects changes its behaviour and its representational form according to these partonomic relationships. For instance a major road might be modelled in a different way if its part of a city (servicing the daily commute) as compared to its role in a rural setting – in which the road more serves to connect cities. These different behaviours result in different cartographic visualisations. Thus we need to determine which section of the road is part of the city, and which part is part of the rural scene. This process creates a new set of hierarchies which is called an aggregation hierarchy (Figure 5) and the relationship it signifies is called a partonomic relation (Mackaness and Edwards 2002; Peng 1997). Aggregation hierarchies specify relations between higher order classes and lower order objects and classes (Liu et al. 2003). It may make use of the classification hierarchy. Thus the aggregation hierarchy tells us how to aggregate the objects belonging to different classification hierarchies to create the objects required in the target database.

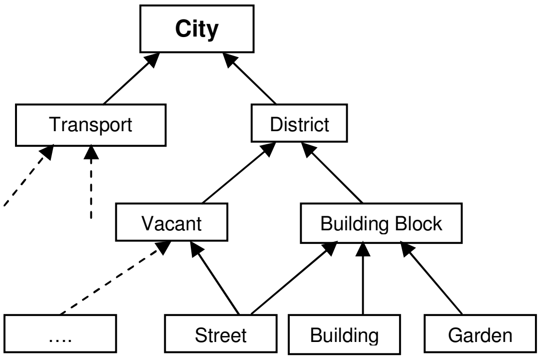

Figure 5: Example of an Aggregation Hierarchy

It is important to note that partonomic relations are application and level of detail dependent. Also that these partonomic structures may not be mutually exclusive. These partonomies may exist as hierarchical structures – palimpsest like. It is necessary to define objects as 'empty containers requiring a particular set of ingredients'. Once defined, we can search the database at the fine scale for that particular set of ingredients (objects), and thus identify an occurrence of the higher order object (akin to searching for a pattern or signature within a dataset). As these higher order objects become instantiated, their behaviours can be associated with other objects. For example, once a number of cities have been identified, we can cross link this information with objects that traverse that city boundary. For example a railway network could be attributed according to whether it was a rural, suburban or city railway. New relationships can be formed between these higher order objects.

3.0 Design and Implementation

An implementation was carried out for the derivation of a synoptic database (1:250,000) directly from a large scale database (OS MasterMap). The platform selected for the implementation was Java, SQLJ and Oracle 10g. The spatial Java Geometry API provided by Oracle offers a wide range of functions to manipulate spatial objects in Oracle (Oracle 2005). Oracle 10g supports all the geometrical and topological functions defined by OGC (Oosterom, Quak, and Tijssen 2002). The overall design of the proposed system is summarised in Figure 6. The following sections discuss the important stages of the implementation.

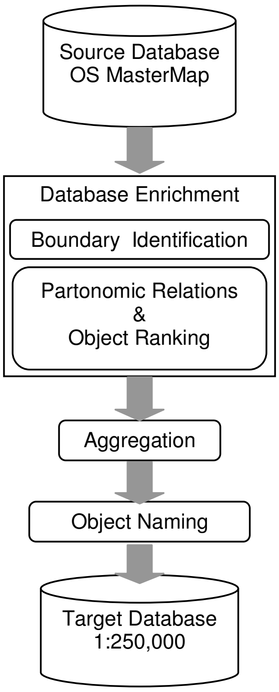

Figure 6: Overall design of the system

3.1 Boundary Identification

Database transformation takes place when a new data model is introduced. The data model defines the classes and instances of the classes for the required application. In the current implementation our new data model had four main classes - Settlement, Woodland, General Surfaces and Roads. To create objects of these classes from the objects of the source database we first need to define their taxonomic and partonomic relations. Objects in the source database (OS MasterMap) are classified into classes such as Buildings, Land, Coniferous tress, non-Coniferous trees, Scrubs, Minor roads, A road, B Road. In order to create objects of new classes of the target database we need in addition to this classification the partonomic relations of each object in the source database.

A clustering algorithm was developed which used 'signature' or typical objects (Mark, Smith, and Tversky 1999) of each class and created a polygon around those objects. By 'signature object' we mean a class in the target database has characteristically defines the higher order object at the target scale. For instance for a settlement class the defining objects are the buildings, even though a settlement includes other classes such as roads, railways, trees. A settlement is identified by a high density of buildings just as a forest region is identified by a high density of trees. This clustering algorithm made use of proximity between the key objects for each target class and returned a polygon around these objects (Figure 7). The area of returned polygon was compared to a threshold size for removal of smaller polygons (Chaudhry and Mackaness 2005).

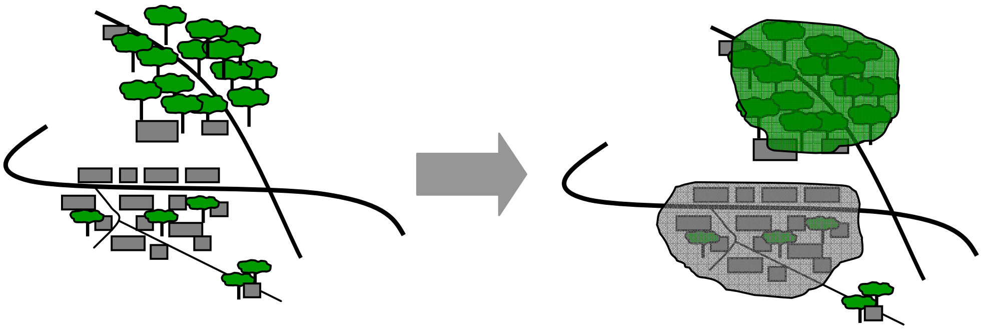

Figure 7: Creating boundaries around dense clusters of buildings

3.2 Database Enrichment

Once the boundary for each class has been determined the next step was to determine the partonomic relations for all objects (not just the signature object) in the source database. Using the boundaries and topological function ('overlap') all objects that are contained by the boundary polygon of the target database classes were identified (Figure 7). For instance in Figure 7 objects that lie completely within the Settlement boundary are reclassified as objects of the Settlement Class. Similarly the same process is applied to objects that lie within the Forest Object boundary.

Some objects lie at the edge of the boundaries, or transit the boundaries (Figure 7). For areal objects, we calculated the percentage of their area within the boundary and a threshold used to determine its membership. Different parts of the network may have different partonomic forms – parts belonging to the city, other parts belonging to rural partitions, yet at all times it is necessary to maintain the integrity and connectivity of the network. Roads were modelled using the Oracle Network Model which enables modelling of the network as a set of edges and nodes and ensures connectivity of the network during generalisation. For all those edges whose deletion or aggregation will create a disconnected network were given an object ranking of 'High'. Similarly for all those objects that we don't want to be eliminated through deletion or aggregation was given a 'high' object ranking. OS MasterMap's ITN (Integrated Transport Network) layer was used to populate the Network Model tables. This resulted in an enriched database in which each object is reclassified into a target class, a target object together with its importance in terms of object ranking. A sample of the database table after the process of enrichment is shown in Table1. It's important to point out that only a section of the attributes is shown here. The 'toid' is the unique identifier of each object. 'Initial class' defines how the object was classified at OS MasterMap, and 'Target Class' defines its new class i.e. new taxonomy, 'Target Object' defines the partonomic relation for the target database object and 'Object Rank' defines the importance of the object at the target database level.

Table 1: Example of an Enriched database with Target Object (part of), Target Class (Taxonomy) and object ranking. Highlighted fields will be aggregated to create a Settlement1 object

Toid	Initial Class	Target Object	Target Class	Object Rank
100000421564	Building	Settlement1	Settlement	Low
100000421565	Building	Settlement1	Settlement	Low
100000421566	Building	Settlement2	Settlement	Low
100000421567	Building	Settlement2	Settlement	Low
100000421568	Coniferous Trees	Forest1	Woodland	Low
100000425169	Coniferous Trees	Forest1	Woodland	Low
100000425170	Coniferous Trees	Settlement1	Woodland	Low
100000425171	Minor Road	Settlement1	Roads	High
100000425177	A Road	Settlement1	Roads	High
100000425178	Motorway	Settlement1	Roads	High
100000425181	Street	Settlement1	Roads	Low

3.3 Aggregation

After the database enrichment stage each object in the database has been reclassified into a resultant class in the target database. In order to create objects of the target database classes we implemented an aggregation algorithm. The algorithm uses the new classification and object ranking to create higher order objects. In Table 1 for instance to create a Settlement1 object all objects that are part of ‘Settlement1’, and object ranking ‘low’ will be aggregated to create object Settlement1 and its class will be ‘Settlement’. Similarly to create a object of Forest1 all objects that are part of the same object of woodland and ‘low’ object ranking will be aggregated to create the resulting object.

3.4 Object Naming

Once the objects of the target database have been created an algorithm was developed to label each object (giving meaningful names to the newly created objects rather than Settlement1, Settlement2, etc.). Labels were drawn from Ordnance Survey’s Strategi database (1:250,000). This database contains an annotation layer in which names are stored as point objects, together with an x, y coordinate. The naming algorithm calculated the nearest point objects from the OS Strategi for each target object to be named. This was done to limit the search for possible labels; only those point objects were selected that were within a certain threshold distance of the object. The algorithm was flexible on assigning more then one label to an object.

4. Case Study and Evaluation:

The proposed deign was implemented in Oracle 10g and Java. Figure 8 shows an input area which consists of 25,000 objects that belong to more then 10 distinct classes. Figure 9 shows the corresponding output in which all object have been reclassified and now belong to one of four classes (Settlement, Woodland, General surface and Roads). It is important to stress that

Figure 9 is just the visual output of the generalised database – it is not considered a cartographic solution. It demonstrates the meaningful aggregation of source objects into higher objects of the target database based on proximity, classification hierarchy and partonomic relations.

Figure 8: Input Dataset OS MasterMap. Enlarged area shows the level of detail (Copyright OS MasterMap)

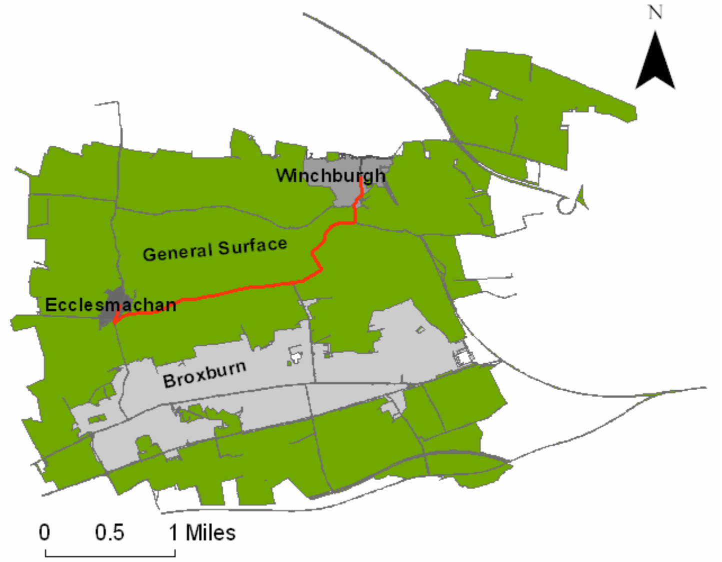

Figure 9: Output at 1:250,000. The Red line shows the shortest road network

4.1 Utility

The new database is useful both for cartographic purposes and for spatial analysis. The output shown in Figure 9 can be used as an input to a cartographic generalisation system (processes such as displacement, enlargement, and simplification can be applied to make the output more visually optimal).

The utility of an enriched database lies in its ability to support spatial analysis routines that were not possible using the source database. For instance if a client is interested in finding all houses that are contained by a city, it's not possible to answer this question in the original database (OS MasterMap) since there isn't a concept of 'city' in it. But a simple SQL (shown below) can be performed on our enriched database.

```
SELECT    a.ID
FROM      ENRICH_DATABASE a, OUTPUT_DATABASE b
WHERE     a.PARTOF=b.PARTOF AND B.OBJECT_NAME='BROXBURN';
```

In this example we are able to do a mapping between a 'city' and 'building' objects in the database because we have explicitly defined the relationship between the two. This has been possible by defining the partonomic relation between the new concept and source objects. Although we have a classification of building objects in the source database these objects have no information about their target class thus this query cannot be performed on the source database. More sophisticated analysis such as finding the shortest road network between the cities can also be performed on the output database. For this we implemented an algorithm that calculated the shortest road network between two given towns. The highlighted road in Figure 9 is a result of one such query which gives the shortest path between the towns of 'Ecclesmachan' and 'Winchburgh'.

5. Conclusions

Generalisation is a process of transforming information from one form to another in order to reveal different qualities and characteristics of the phenomena being modelled. In this paper we have argued that generalisation is not limited to the cartographic domain. Once a database has been generalised it can then be used for different spatial analysis routines or as an input to cartographic generalisation. Further work will look into extending our classification (hills, mountain ranges, etc) and finding their partonomic relations. Future work will also look into dealing with fuzzy membership.

Acknowledgements:

We gratefully acknowledge support and funding from the Ordnance Survey and The University of Edinburgh.

References:

Brassel, K E, and R Weibel. 1988. A review and conceptual framework of automated map generalization. *International Journal of Geographical Information Systems* 2 (3):229-244.

Chaudhry, O, and W Mackaness. 2005. Visualisation of Settlements Over Large Changes In Scale. Paper read at International Cartographic Association Generalisation Workshop, 7-8 July, at La Coruna.

Downs, T C, and W A Mackaness. 2002. Automating the Generalisation of Geological Maps: The Need for an Integrated Approach. *The Cartographic Journal* 39 (2):137-152.

Foerstner, W, and L Pluemer. 1997. Semantic Modelling for the Acquistion of Topographic Information from Images and Maps. Paper read at Proceedings of the Workshop “SMATI ’97, 21-23 May, at Birkhäuser Verlag.

Goodchild, M, and J Proctor. 1997. Scale in a Digital Geographic World. *Geographical and Environmental Modelling* 1:5-23.

João, E M. 1998. *Causes and Consequences of Map Generalisation*: Taylor & Francis Ltd.

Kilpelainen, T. 1997. *Multiple representation and generalisation of geo-databases for topographic maps*. Vol. No 124: Publications of the Finnish Geodetic Institute.

Kilpelainen, T, and T Sajakoski. 1995. Incremental generalisation for multiple representations of geographical objects. In *GIS and Generalisation: Methodology and Practice*, edited by J. C. Muller, J. P. Lagrange and R. Weibel.

Levin, S A. 1992. The Problem of Pattern and Scale in Ecology. *Ecology* 73 (6):1943-1965.

Liu, Y., M. Molenaar, T Ai, and Y Liu. 2003. Categorical Database Generalisation Aided by Data Model. Paper read at Proceedings of the 21st International Cartographic Conference ICC, 10-16 August, at Durban, South Africa 10-16 August 2003.

Mackaness, W, and G Edwards. 2002. The importance of Modelling Pattern and Structure in Automated Map Generalisation. Paper read at Joint Workshop on Multiscale Representation of Spatial Data, at Canada.

Mark, D. 1999. Spatial Representation: A Cognitive View. In *Geographical Information Systems: Principles and Application*, edited by Maguire, Goodchild, Rhind and Longley.

Mark, D. M., B. Smith, and B. Tversky. 1999. Ontology and Geographic Objects: An Empirical Study of Cognitive Categorization. In *Spatial Information Theory: A Theoretical Basis for GIS, Berlin: Springer-Verlag, Lecture Notes in Computer Science*, edited by C. Freksa and D. M. Mark.

Molenaar, M. 1998. *An Introduction to the Theory of Spatial Object Modelling for GIS*. London: Taylor and Francis.

Molenaar, M. 2004. Spatial Data Modelling, Dual Partitions and the Specification of Semantics. Paper read at XXth ISPRS Congress, 12-23 July 2004, at Istanbul, Turkey.

Molenaar, M, and D E Richardson. 1994. Object Hierarchies for Linking aggregation levels in GIS. Paper read at Proceedings of the Symposium of ISPRS Commission IV, at Athens, Georgia, USA.

Monmonier, M.S. 1984. Geographic information and cartography. *Progress in Human Geography 8*:381-391.

Mustiere, S, and B Moulin. 2002. What is Spatial Context in Cartographic Generalisation. Paper read at Symposium on Geospatial Theory, Processing and Application, at Ottawa, Canada.

Oosterom, P.J.M., C.W. Quak, and T.P.M. Tijssen. 2002. Testing current DBMS products with real spatial data. Paper read at In Proceedings UDMS 2002, at Prague, Czech Republic.

Oracle. *Spatial User's Guide and Reference 10g Release 1 (10.1)* 2005 [cited 15th Feb 2005. Available from http://www.oracle-10g-buch.de/oracle_10g_documentation/appdev.101/b10826/toc.htm.

Ormsby, D., and W. A Mackaness. 1999. The Development of Phenomenon logical Generalisation Within an Object Oriented Paradigm. *Cartography and Geographical Information Systems* 26:70-80.

Peng, W. 1997. *Automated Generalization in GIS*. Edited by ITC. Enschede: ITC.

Smaalen, J W N van. 2003. Automated Aggregation of Geographic Objects: A New Approach to the Conceptual Generalisation of Geographic Databases. *Doctoral Dissertation, Wageningen University, The Netherlands*.

Smith, J M, and D C P Smith. 1977. Database abstractions: aggregation and generalisation. *ACM Transactions on Database Systems* 2:105-133.

Thompson, P J. 1989. *Data with semantics: Data models and data management*. New York: Van Nostrand Reinhold.

Turnbull, David. 1989. *Maps are territories Science is an atlas*: Deakin University.

Weibel, R, and G Dutton. 1999. Generalising Spatial Data and Dealing with Multiple Representations. In *Geographical Information Systems*, edited by P. Longley, M. F. Goodchild, D. J. Maguire and D. W. Rhind. New York: John Wiley.

Worboys, M. 1996. Metrics and topologies for geographic space. Paper read at Advances in GIS Research II: proc of 7th International Symposium on Spatial Data Handling.