A Methodology for Defining and Inferring Higher Level Semantic Information from Topographic Data

Marie-Kristina Thomson

Department of Geomatic Engineering, University College London, Gower Street, London WC1E 6BT, United Kingdom
Telephone: +44 (0)20 7679 2740
thomson@ge.ucl.ac.uk

Abstract. Lack of high level semantic descriptions within geospatial data has been a major critique point for the handling and usage of such data. Much research has been invested into the cognitive principles of how humans conceptualise, store and work with spatial information, yet current data models continue to fail to meet the users' needs for flexible and fit-for-use data. This paper proposes a general methodology for defining and inferring higher level semantic information in the form of functional attributes from topographic data by the means of ontologies. Based on the argument that a specific spatial structure and configuration is inherent in different land use types, this project focuses on the ontology's reasoning capabilities for making the modelled functional information – from both the human and data point of view – explicit within the data, thus enriching its current thematic contents with higher level concepts and enabling new opportunities for on demand mapping and multi-representation.

1. Introduction

Semantic heterogeneity is a consistent issue among geospatial data sources prohibiting unrestrained data sharing and integration as well as hindering human interaction with such data – let alone the data's insufficiency to meet specific requirements and task scenarios due to lack of higher-level semantic information. Thus, making geographical data available to specific types of users and the problems they need to solve is of primary concern to data providers who are driven by demand. One way to semantically annotate geographical data with higher-level concepts is by explicating this knowledge by means of ontologies, i.e. explicitly formalising specification of shared conceptualisations (Klien and Lutz, 2005). This project investigates how this can be achieved for functional information by tracing how functional top-level concepts such as residential link down to their underlying individual geographical entities and their relations, which are the constituting building bricks that make up these high-level representations of the land use.

On the one hand, ontologies can be applied for making the semantics of the information content of geospatial data explicit in order to enhance geographic information and to allow better interaction between the user and the data. On the other hand, ontologies provide a means to discover and retrieve implicit information from geospatial databases. Logical reasoning inherent in ontologies can be used to discover implicit relationships between human concepts and information descriptions within the data, as well as to flexibly construct taxonomies for classifying information sources (Lutz and Klien, 2006).

The most important concern that this research addresses is how to bridge the gap between the widely deployed models of space and what research in cognitive science and related fields identified as being important for human interaction with and conceptualisation of space (Mark *et al.*, 1999). The first step towards achieving this goal is to make higher-level semantic information automatically available from existing data repositories, thus providing ease with the handling and usage of such data for a variety of purposes.

2. Current Issues

Conventional GIS data models suffer shortcomings in the way geographic information is stored and represented. It has been widely acknowledged that such representations fail to meet specific application contexts, as it is yet impossible to acquire and store all knowledge from raw information before knowledge is accessed, and providing unprocessed raw information and computing specific knowledge on demand (Freksa and Barkowsky, 1995; Peuquet, 1988; Burrough and Frank, 1995; Frank, 1992). These deficiencies mainly arise due to lack of incorporating any explicit consideration of how humans cognitively store and use geographic knowledge. These issues are widespread, and affect both the direct use of data as well as their integration with other data sources. The question that needs to be answered is how higher-level knowledge can be derived from such data and become part of the existing data model.

3. Proposed Solution

Higher-level knowledge is crucial for the fitness of use of geographic data, but so far has only resided in the head of the user, or in separately defined models (Mennis *et al.*, 1999). As illustrated in Frank and Mark (1991), GIS research must separate the concepts involved in a programme from the mechanics of its implementation as a programme, hence separating the conceptual database schema from the physical storage arrangement. A third schema describes subsets of the conceptual view according to users and their specific task contexts, as illustrated in Figure 3.1. According to these schema representations, this research addresses the following questions: how do people perceive and conceptualise land use information? How can we formalise land use conceptualisations and relate them to the data? How is land use information stored (explicitly as well as implicitly) within the data's objectified real world representation? And how can we bridge the gap between these three representations and incorporate the higher level semantics as described by the users' view into existing data sources?

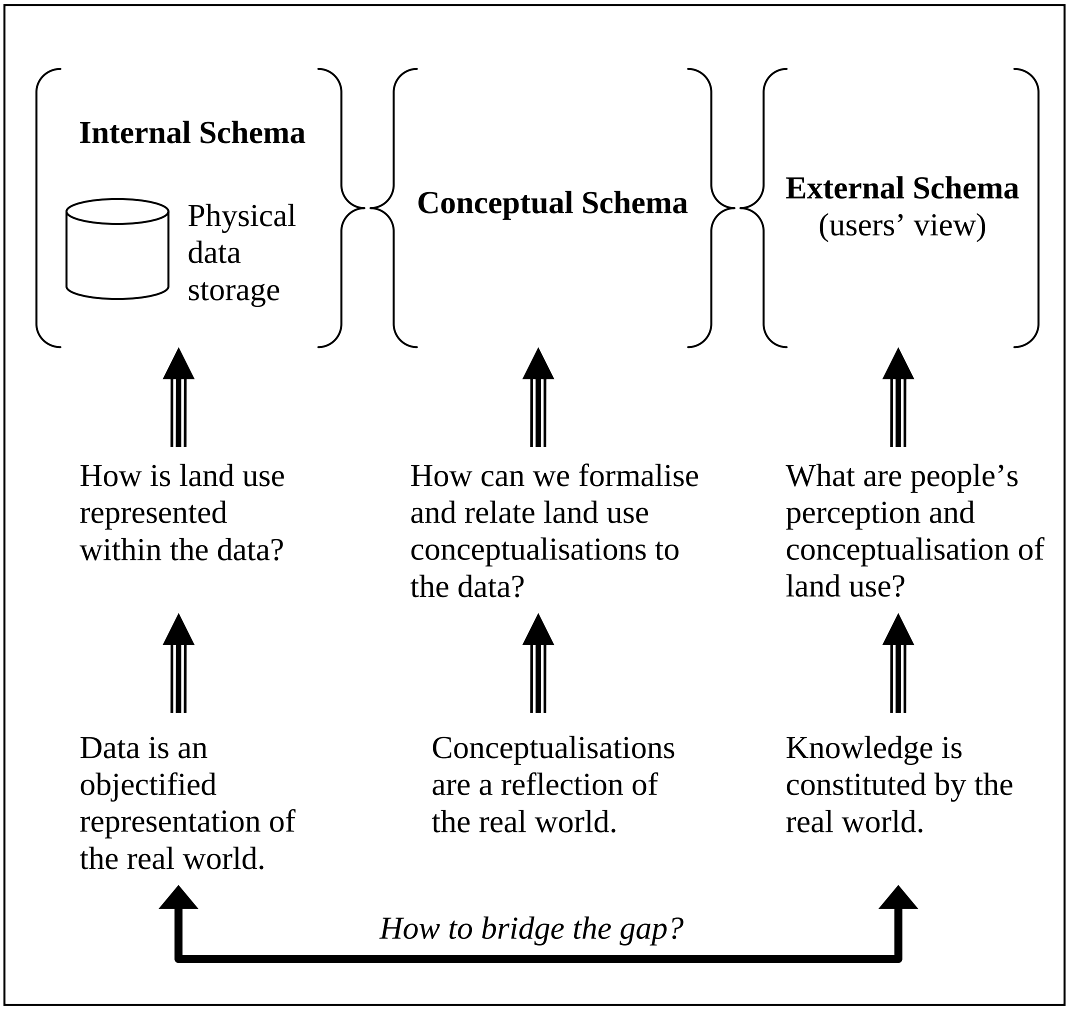

Figure 3.1 Concerns of this research regarding the inference and integration of higher level semantics, as described by the users' schema, with the data's internal schema.

Since we are dealing with a domain that manifests itself with its underlying geographic existentialism, higher level knowledge of land use can be grounded to its finest level of detail in terms of objects, attributes, and relations that constitute its superordinate levels of information. For example, the high level semantics of *residential* can be related to the landscape through its make up of geographic objects, their affordances and how they relate to one another in order to allow the use of that geographic space for human habitation. The survey in Thomson (2007) collected conceptualisations that relate land use types to the landscape from the human (or user's) perspective, as illustrated in Figure 3.2. These conceptualisations are a specification of different spaces that define and relate land use to its underlying topography – from the semantic categories with which people communicate about a given domain, their defined relational, functional and attribute spaces, to their underlying real world object space. The complexity of the geographic domain can only be captured by accounting for all these aspects whose relations and affordances determine how the space can be used.

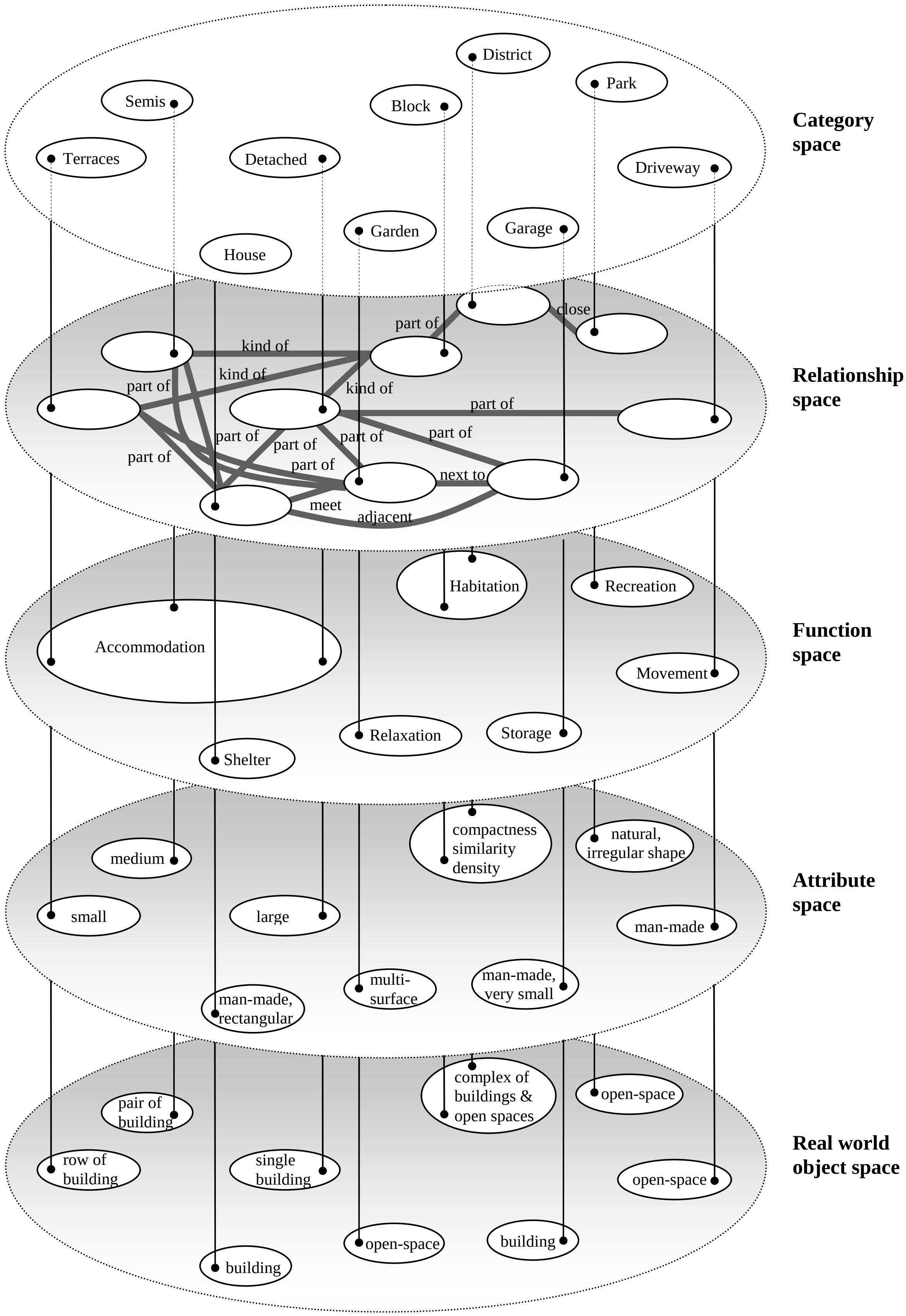

Figure 3.2 Conceptualisation of Residential

In addition, the survey revealed the respondent's success in interpreting plain topographic maps according to land uses due to their ability to draw analogies from the known to the unknown. For example, the external schema, as described in Figure 3.1, addresses the respondent's cognitive use of *a priori* knowledge to interpret and categorize new observational data, i.e. topographic maps. The schema is a set of information about a type of object or category that is used to discover new instances of this type of object or category. Hence, a respondent's schema for the visual recognition of a 'residential' pattern may include information such as: small objects in a row running adjacent to a road, or pair of houses next to garden spaces running adjacent to a road. Thus, when the respondent recognises objects that meet these schema criteria, he or she can identify the geographic space in the map as an instance of the residential category with specific values for the generic properties described by the schema; such that buildings will have a certain size, a certain shape, and a certain alignment. As summarised by Mennis *et al.* (2000, p.504) "*schema are formed by induction from repeated experience with the same type of object and may be based on the prototype example of a category of objects. A schema is not an exact representation, but is more like a general pattern.*"

In order to translate this ability into mechanised ways, such *a priori* knowledge of a general representation of the domain of interest is required. This can be achieved by defining and formalising higher level knowledge as ontologies and linking them to the database. Since the domain ontology can be grounded onto its underlying geography, it becomes possible to map the domain ontology onto the data for classifying the data's features according to the domain ontology's specification, thus allowing instances of the data source to inherit the defined higher level semantic information. Figure 3.3. illustrates this proposed solution. A domain ontology of higher level semantic information is derived from the external schema, i.e. users. A data ontology is developed making the semantics embedded in the underlying geographic data model's primitives explicit, i.e. the internal schema, which is based on an object data model for representing topographic space that structures the world by discrete, identifiable entities, with a geometrical representation and descriptive attributes. By combining these two approaches into a shared conceptualisation of the application ontology, i.e. into the conceptual schema, it should become possible to map the conceptualisation to the data and search *a priori* for features that meet the conceptualisation's definition of higher level functional information. For example, having various levels of semantic definitions for the category *residential* allows to express the category at a high level as districts and blocks, and at finer levels as terraces – semis – detached, and row of buildings – pair of buildings – single buildings, and so forth. The high level semantic description can then be inherited down to the finest level, e.g. that of individual objects that meet the schema's criteria of a certain size, shape, type, etc. These criteria are determined by both the knowing *what* an object is, in terms of detailed and precise geometric properties, such as shape, relative orientation of component parts, size, etc., and knowing *where* something is, in terms of locational properties of objects such as containment, distance and direction. All these aspects are embedded in a knowledge of environmental scenes and knowing what human activity takes place on them. As a result, the data's thematic contents become richer and reflect better the need for higher level information relevant to the end user ensuring compatibility with possibly other sources of data. Finally, by separating the high level domain ontology from the more detailed data ontology, the sharing and applicability of the domain ontology is ensured, as cultural variations in terms of different spatial configurations that exhibit the same land use categories are captured within the respective, local data ontologies and not the higher level semantic, global domain ontologies.

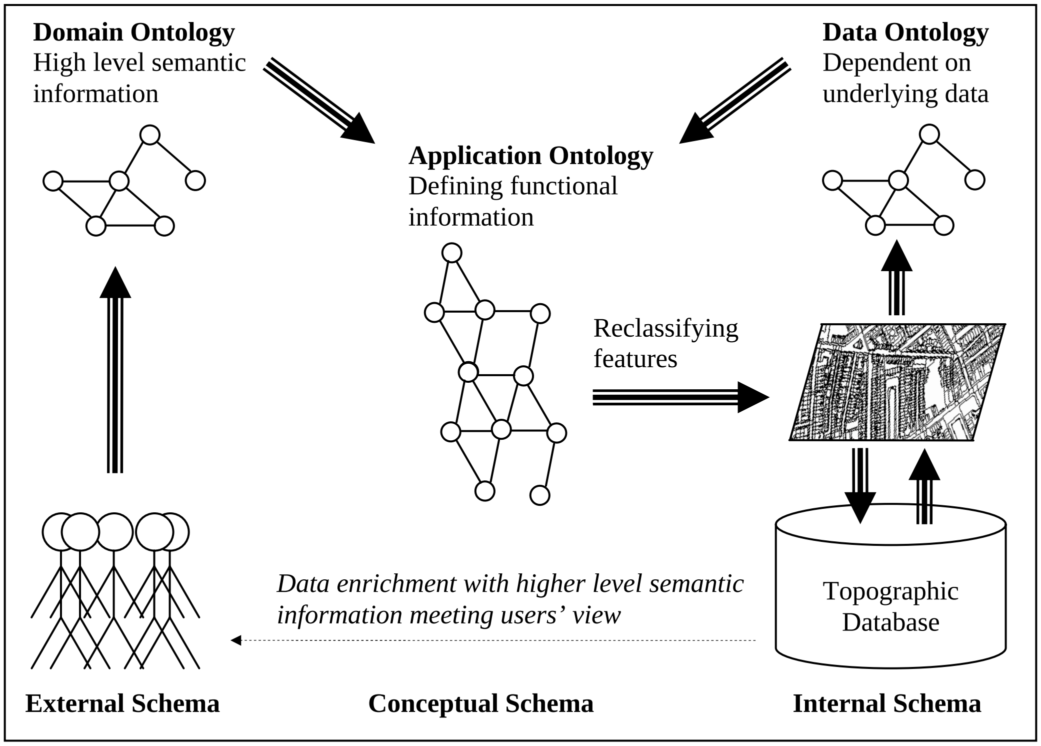

Figure 3.3 Proposed methodology

The need for integrating semantics into database representations is not new, and research has been ongoing since the advent of the relational data model (Codd, 1970; Chen, 1976, Hull and King, 1987; Peckham and Maryanski 1988; Borgida, 1991; Booch, 1994). Indeed, there have been many studies and proposals on how to incorporate cognitive principles into geographic database representations, for example Mennis *et al*. (2000) proposed a pyramid framework decomposing space, time and theme properties so as to model the where, when, and what components of people's knowledge representations allowing to build knowledge hierarchies describing multiple structural interpretations of observed geographic data. Fonseca *et al.* (2003) proposed a way to link the formal representation of semantics, i.e. ontologies, to conceptual schemas describing information stored in databases illustrating a formal framework that explains a mapping between a spatial ontology and a geographic conceptual schema.

Here, the aim is to make high-level concepts explicit within the information stored in databases through an ontological reasoning system. The resulting semantic enrichment allows both the users to access these information based on top-level semantics as well as to classify topographic data according to how the geographic space they represent is consumed by human activity at a coarse level. Guarino (1998) coined the term ontology-driven information systems for such systems that make use of formally defined ontologies. However, the current theory on the use of ontologies addresses the broader intention of providing a basis for knowledge consolidation and exchange lacking of practice in the use of ontologies for real-world problem solving along with the scarcity of consistent ontologies, hence going far beyond the capabilities of current data modelling tools and techniques (Fonseca *et al*., 2003). Data sharing has been indeed the main motivation for ontology development (e.g. Fonseca, 2001), and is often a resulting characteristic of formally defined ontologies. In this case, however, the focus is on reasoning with ontologies in order to derive implicitly stored information from topographic data as well as annotating such information with higher level semantics. For example, data on geographic space are primarily collected as individual entities (the object model) or as continuous spatially varying phenomena (field view) for the purpose of describing the topography; however, the user or application requires the topography to be represented as a set of areas depicting the human activity that takes place on that geographic space. The aim is to derive these areas – at a coarse level in terms of residential, recreational, industrial, etc. – through the use of some operation that reclassifies the instances within the data according to the specified definitions within the application ontology. Important is that the concepts about the topography or its geographic location characteristics do not change throughout the process of deriving an alternative representation, but become merely enriched through higher level semantic information and top-level concepts. This is *a posteriori* solution to the problem of providing a high-level view of the real world-representation in existing data repositories, as opposed to the suggestion in Fonseca *et al*. (2003) of integrating ontologies into the conceptual representation design of implementing the physical database design.

4. Application

This project focuses on Ordnance Survey's (OS) national topography product for Great Britain, hence aiming to add geographic meaning against OS cartographic features. From their business perspective this project has potential to contribute towards an open architecture for providing automatic on-demand mapping capabilities looking at the wider problem of generating user-specified maps with different contents and scales (Regnauld, 2006). OS itself is interested in developing a domain ontology that represents the geographic and topographic domain from the OS view of the world, while at the lower level a data ontology is supposed to link the database to the domain ontology. Therefore, not only the developed ontology in this project, but also the enrichment of OS data with higher level semantic information increases opportunity for semantic operability. In terms of allowing map users to reason about the geography at multiple levels of abstraction, Fonseca *et al.* (2003) mention that kinds of relationships commonly found in object-oriented conceptual schemas, such as generalisation, specialisation, and aggregation, can be mapped to semantic relationships in an ontology. For example, Edwardes *et al.* (2005) argue that map generalisation can be seen as facilitating an exchange of knowledge between the map producer and the map user by constructing a particular view of the world, for which ontologies can provide the mechanism by which the concepts inherent in this knowledge can be explicitly represented. Abstract geographic entities that will be present in the generalised map, as well as their inter-relationships and logical entailments, can be defined with an ontology, thus allowing generalisation operations to be performed and constraints described using a single logical model. Kulik *et al*. (2005) also examined ontological information with respect to generalisation, as the conformance with particular requirements of a user engaged in a specific task requires that information retrieval and processing operations must be able to incorporate ontological information about the data, the user, and the user’s task. Consequently, having such abstract, higher level information about the data’s geographical entities described in this proposed ontology – although in the form of functions – can aid generalisation processes in terms of aggregating those areas that carry the semantic description of different functional information categories at a coarse level, e.g. residential areas at district level, or possibly industrial and commercial areas. According to Regnauld (2001) information conveyed by each object both as a single entity, i.e. the house, and collectively as e.g. the residential area are important for automated generalisation processes. At the disposal is the database describing each object individually and independently from one another, thus representing a single building as a list of coordinates describing its boundary together with a semantic code. However, abstract semantic information that describe objects collectively according to their spatial configuration is not explicitly contained within databases.

5. Future Work

Knowledge for the domain ontology has been collected through questionnaire surveys, which captured people’s general conceptualisations of different land use types in relation to the landscape as well as the concepts that are most commonly used when communicating about land uses (Thomson, 2007). From a professional user perspective the importance of different functional information categories was analysed in Thomson (2006). Future work, therefore, consists of establishing the data ontology, and combining and modelling the different levels of knowledge within the application ontology. This consists of analysing how well land use is represented within the database by investigating relevant features within OS MasterMap according to geometrical, thematic and shape attributes. Measures for making such variables (e.g. size, shape, topology, etc.) explicit in terms of their parametric thresholds are summarised in Steiniger and Weibel (2005). As described in the previous sections, the purpose of the data ontology is to bridge the gap between the domain ontology and the database schema by describing the schema plus the necessary implicit information and the map (Hart, 2007). Formal concept analysis (e.g. Kavouras and Kokla, 2002) may be useful in order to analyse commonalities and overlaps between categories and assigning relevant attributes. With respect to the implementation, the most suitable tools, implementation languages, and platforms need to be analysed and justified by acknowledging limitations of the existing conceptual models such as category-based ontologies, conceptual spaces, affordance based models, image schemata, and multi representation (Tanasescu, 2007). It still needs to be clarified how the application ontology can be linked to the database, especially with respect to previous attempts that identified limitations in terms of modelling spatial relations (Hart, 2007), and how the formalised knowledge can be best applied for classifying the data’s features, keeping in mind that the expected outcome remains questionable in terms of coarseness as well as accuracy.

6. Conclusion

Issues with existing geospatial data models were highlighted in terms of their lacking correspondence with people's cognitive abilities and conceptualisations of space. Problems with semantic heterogeneity of data sources together with their limited abstract, high-level information contents continue to hamper the way people handle and integrate geographic data. Thus, the need for content-richer data with more meaningful descriptions is still apparent, despite research efforts for integrating semantics into database representations that have been ongoing since the advent of the relational data model. A methodology is presented here that makes use of ontologies for formalising knowledge about shared land use conceptualisations from their high-level abstract semantics down to the constituting individual real world objects and relations that make up the physical reality of the way human make use of the environment. It is to be studied whether a formalised conceptualisation of the required information – derived from both the underlying data as well as the human cognitive representation – can serve as a sufficient knowledge base for reasoning about instances within a database for classifying its features according to higher-level semantics as defined within a conceptual schema of the application ontology.

7. References

Booch, G. (1994) *Object oriented analysis and design*. New York: Benjamin/Cummings Publishing Co.

Borgida, A. (1991) Knowledge representation, semantic modelling: similarities and differences. In Kangassollo, H. (Ed): *Entity-Relationship Approach: The Core of Conceptual Modelling*. North-Holland: Elsevier Science Publication, pp. 1-24

Burrough, P. A. and Frank, A. U. (1995) Concepts and paradigms in spatial information: are current geographical information systems truly generic? *International Journal of Geographical Information Systems* **9**(2): 101-116

Chen, P. P.-S. (1976) The entity-relationship model – toward a unified view of data. *ACM Transactions on Database Systems* **1**(1): 9-36

Codd, E. F. (1970) A relational model of data for large shared data banks. *Communications of the ACM* **13**: 377-387

Edwardes, A., Burhardt, D. and Neun, M. (2005) Interoperability in Map Generalisation Research. In *Proceedings of the International Symposium on Generalisation of Information, ISGI 2005*

Fonseca, F. (2001) *Ontology-driven geographic information systems. In Spatial Information Science and Engineering*, PhD Thesis, University of Maine: Orono

Fonseca, F., Davis, C. and Câmara, G. (2003) Bridging ontologies and conceptual schemas in geographic information integration. *GeoInformatica* **7**(4): 355-378

Frank, A.U. (1992) Spatial concepts, geometric data models, and geometric data structures. *Computers & Geosciences* **18**(4): 409-417

Frank, A. U. and Mark, D. M. (1991) Language issues for GIS. In Longley, P.A., Goodchild, M.F., Magcuire, D. J. and Rhind, D. W. (Eds): *Geographic Information Systems*, 1st ed., Canada: John Wiley & Sons, pp.147-163

Freksa, C. and Barkowsky, T. (1995) On the relation between spatial concepts and geographic objects. In *the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends*. Intelligent Systems, pp.26-34

Guarino, N. (1998) Formal ontology and information systems. In Guarino, N. (Ed): *Formal Ontology in Information Systems*, IOS Press: Amsterdam, The Netherlands, pp.3-15

Hart, G. (2007) *Ontology Database Connectivity – Field Definition.* Ordnance Survey Research Labs, Internal report, Crown copyright

Hull, R. and King, R. (1987) Semantic database modelling: survey, applications and research issues. *ACM Computer Surveys* **19**(3): 201-260

Kavouras, M. and Kokla, M. (2002) A method for the formalization and integration of geographical categorizations. *International Journal of Geographical Information Science* **16**(5): 439-453

Klien, E. and Lutz, M. (2005) The role of spatial relations in automating the semantic annotation of geodata. In Cohn, A. G. and Mark, D. M. (Eds.): *COSIT 2005*, LNCS 3693, Springer-Verlag Berlin Heidelberg, pp. 133-148

Kulik, L., Duckham, M. and Egenhofer, M. (2005) Ontology-driven map generalization. *Journal of Visual Languages and Computing* **16**: 245-267

Lutz, M. and Klien, E. (2006) Ontology-based retrieval of geographic information. *International Journal of Geographical Information Science* **20**(3): 233-260

Mark, D. M., Freksa, C., Hirtle, S.C., Lloyd, R. and Tversky, B. (1999) Cognitive models of geographical space. *International Journal of Geographical Information Science* **13**(8):747-774

Mennis, J. L., Peuquet, D. J. and Qian, L. (1999) A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* **14**(6): 501-520

Peckham, J. and Maryanski, F. (1988) Semantic data models. *ACM Computing Surveys* **20**(3): 153-189

Peuquet, D. J. (1988) Representations of geographic space: toward a conceptual synthesis. *Annals of the Association of American Geographers* **78**(3): 375-394

Regnauld, N. (2001) Contextual Building Typification in Automated Map Generalisation. *Algorithmica* 30: 312-333

Regnauld, N. (2006) *Open architecture to provide on demand mapping.* Southampton: Ordnance Survey, Internal report.

Steiniger, S. and Weibel, R. (2005) Relations and structures in categorical maps. In: *8th ICA WORKSHOP on Generalisation and Multiple Representation*, A Coruña, Spain.

Tanasescu, V. (2007) Spatial semantics factoring with differences. To appear in COSIT’07, Melbourne, Australia

Thomson, M.-K. (2007) *Human knowledge acquisition of functional information*. Internal Report, London: UCL, Department of Geomatic Engineering.

Thomson, M.-K. (2006) *Analysis of functional information – needs and opportunity analysis, user requirements, and use case.* Internal Report, London: UCL, Department of Geomatic Engineering.