Dwelling *on* Ontology –

Semantic Reasoning over Topographic Maps

Marie-Kristina Thomson

Thesis submitted for the Degree of
Doctor of Philosophy (PhD)

UCL

Declaration of Originality

I, Marie-Kristina Thomson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

…………………………………….. 2009

Abstract

The thesis builds upon the hypothesis that the spatial arrangement of topographic features, such as buildings, roads and other land cover parcels, indicates how land is used. The aim is to make this kind of high-level semantic information explicit within topographic data. There is an increasing need to share and use data for a wider range of purposes, and to make data more definitive, intelligent and accessible. Unfortunately, we still encounter a gap between low-level data representations and high-level concepts that typify human qualitative spatial reasoning. The thesis adopts an ontological approach to bridge this gap and to derive functional information by using standard reasoning mechanisms offered by logic-based knowledge representation formalisms. It formulates a framework for the processes involved in interpreting land use information from topographic maps. Land use is a high-level abstract concept, but it is also an observable fact intimately tied to geography. By decomposing this relationship, the thesis correlates a one-to-one mapping between high-level conceptualisations established from human knowledge and real world entities represented in the data. Based on a middle-out approach, it develops a conceptual model that incrementally links different levels of detail, and thereby derives coarser, more meaningful descriptions from more detailed ones. The thesis verifies its proposed ideas by implementing an ontology describing the land use 'residential area' in the ontology editor Protégé. By asserting knowledge about high-level concepts such as types of dwellings, urban blocks and residential districts as well as individuals that link directly to topographic features stored in the database, the reasoner successfully infers instances of the defined classes. Despite current technological limitations, ontologies are a promising way forward in the manner we handle and integrate geographic data, especially with respect to how humans conceptualise geographic space.

Table of Contents

List of Figures

List of Tables

Foreword and Acknowledgements

"In GIS, spatial context and general configuration can be used to make explicit some information which otherwise would have remained inaccessible. This project proposes to investigate the potentialities of such an approach to cartography data and map representation in order to extract different information from existing databases, i.e. information which is hidden or implicit, and therefore other than what the data were initially harvested for." This brief paragraph is the introductory description of a CASE PhD studentship on cartographic data analysis and enhancement sponsored jointly between the Engineering and Physical Science Research Council (EPSRC) and Ordnance Survey (OS), the national mapping agency of Great Britain. It presents the starting point for this thesis with the assumption that specific land uses have their own specific organisational patterns that lead in turn to specific types of spatial configurations. A database often includes physical descriptions of features, but consistently lacks functional descriptions. The problem addressed here is how such functional information can be derived from OS data. After gathering enough evidence that meaningful functional information can indeed be derived from spatial configurations, because of the inherent relationship between spatial form and function, the quest continues to find a novel approach to exposing these kinds of information. It didn't take long to discover the relevance and importance of spatial reasoning to interpret land uses from topographic map data, both from a human and machine-interpretable point of view. In research, we often look first at ways how people solve a particular problem to translate such skill into a form comprehensible to the machine. If I, as a person, can easily interpret land use information from a topographic map, then it should be possible to access this information in an automated way. But is procedural knowledge encoded in a program flexible enough to achieve such a solution? A paper by Eva Klien and Michael Lutz (2005) raised my curiosity of alternative approaches such as ontology applied in geographic information science and its associated capabilities of logical reasoning and information retrieval. *"John, the user of geospatial web services, is looking for information sources that will answer his question. His query for 'low-lands adjacent to a river that are subject to flooding' is formulated based on a geospatial ontology"* (p.136). Klien and Lutz define *a priori* the spatial constraints of the floodplain as geospatial concepts of their domain ontology. The geometric

characteristics and spatial relations are associated with a spatial analysis method, which can identify and retrieve matching data on the fly. Hence, there is no need for finding a dataset that explicitly stores floodplains, and more importantly this approach promises to offer a more dynamic and cognitive solution to the problem. As a result, I took the route of ontologies, description logics, and cognitive sciences, and discovered the excitement of these fields applied in GIS at recent conferences, their potential benefits and range of applications, but also their limitations and plenty of sceptical views. But how successful is this method in the pursuit to solve my research problem?

Chapter 1

Introduction

"Schools and houses will be close to residential areas. Hospitals will be close to residential areas. Recreational areas will be inside of residential areas. Trains and tubes will link residential areas whilst factories and industries will be outside of or far from residential areas."

–A respondent's view of the residential area

Whenever you take a trip in an airplane and take a look outside the window, given that the weather permits sight, you can see patterns that emerge from the Earth's surface. You can identify agriculturally used parcels of land, forests, and splattered spots of houses and villages, or the dense located buildings in urban areas. The point is that the spatial arrangement and context of features within a landscape can tell you a lot about their purposes and uses. Indeed, many researchers have previously argued that space creates a special relation between function and social meaning, thereby relating spatial configuration to social structure (e.g. Lévi-Strauss, 1963; Hillier and Hanson, 1984).

The notion of a relationship between form and function in urban areas underpins much research in computational urban morphology, despite scepticism about its conceptual basis and potential application to determine urban land use from automated analysis of spatial data (Barr and Barnsley, 2004). Form, in this sense, relates to the urban structure and its manifestation in space. Topographic data, for example, accurately represent the shape of the Earth including the detailed location and morphology of features such as roads and buildings. Function, however, is a much more intricate and difficult concept to ascertain. It defines an activity, depended on socio-economic factors, that is natural to or the purpose of an object. According to Hillier and Hanson (1984), in addition to the practical and social portrayal of its object, function belongs above all to the realm of cultural identity or meaning. This meaning closely relates to land use – a term that refers to the human activity that takes place on, or makes use of, that land (Barnsley *et al.*, 2001). Such activity is intimately linked to what the environment affords, or has to offer. To Gibson (1979), affordances are relationships that point both ways, to the environment and to the observer. Along with cultural and other constraints and

conventions (Norman, 1999), we therefore find ourselves in a setting for our activities that includes schools, homes, the local shopping centre, friends' houses, the pattern of streets and roads, restaurants, cinemas, parking lots, GPs, etc. These are important places that we use, need to know about, and need to represent mentally (Downs and Stea, 1977; Dey and Abowd, 1999).

In this context, it sounds reasonable to posit that the understanding of space is anchored in the experience of people's perception of space, and spatial cognition and behaviour. With the cognitive use of *a priori* knowledge, people can easily interpret and categorize new observational data, and infer new information by induction from repeated experiences. People employ methods of spatial reasoning almost constantly to infer information about their environment (Egenhofer and Mark, 1995). This also holds for the interpretation of geographic representations, such as topographic maps. If we want to bridge the current gap between existing deployed models of space and the way humans cognitively use spatial information, we must design representations that follow human intuition and are, therefore, easily accessible to a large range of users. By relating people's experiences and knowledge about the functioning of the environment to spatial data, we can discover much richer, previously hidden descriptions.

The importance of high-level semantic descriptions is becoming increasingly evident with the need to share and use data for many different purposes, and making data more definitive, intelligent and accessible. Especially to data providers, who are driven by demand and profit margins, a dataset's insufficiency to meet specific requirements and task scenarios due to lack of higher-level semantic information is of primary concern (Lüscher *et al.*, 2007). Often we face disaggregate and heterogeneous data that perhaps depict the same thing but semantically denote different things, prohibiting us to make efficient use of their sources. To match the different requirements that users expect from geospatial information, demands an understanding of the ontological aspects of geospatial data (Kulik *et al.*, 2005). Function, relating to what an object affords, is one of the five basic ontological relations that make geographic information more explicitly meaningful. It is a key characteristic for defining objects in feature-attribute catalogues for geographic information (Rugg *et al.*, 1997). For example, to facilitate on-demand mapping, that is, to extract only relevant information on the fly from a variety of data sources, higher-level entities are required. These need to be readily available as

components, such as urban extent or aggregated buildings, to aid the creation of new, custom and multi-resolution map products. Furthermore, by reflecting better the way people perceive the world, data become conceptually more useful and therefore can be exploited to their full potential (Mennis *et al.*, 2000).

The shortcomings of current data representations inspire disciplines such as knowledge discovery, information retrieval and ontology-driven information systems, and specifically this thesis. We embark on a journey to enrich spatial data conceptually by exposing functional information within topographic data using description logics, and by putting a conceptual model for residential area into operation with off-the-shelf technology. 'Dwelling *on* ontology', one part of the thesis title, relates to the definition and exposure of the dwelling – residence, accommodation, or house, and beyond – through an ontology. We will encounter why much recent research focus seems to dwell on ontology in its current fascination with the apparent 'magic' that ontology offers. The Conference on Spatial Information Theory (COSIT) series, for instance, draws heavily on work in linguistics and cognitive science, and crystallises around the term ontology. However, is ontology the solution to all our problems? We will discover what ontology can realistically offer, and find out if it is worth all this commotion. The second part of the title, 'semantic reasoning over topographic maps', is associated with the processes involved in discovering functional information from topographic maps. This incorporates the way humans differentiate and categorise, make sense of the world, and reason over information, and how we can translate such skills to reason logically within the computational environment. Not only is there the need to incorporate how humans cognitively process and use geographic knowledge, but to make data more intelligible and to provide automated access to implicit information that is contained within.

1.1 Aims and contributions of the thesis

The topography layer of Ordnance Survey MasterMap does not explicitly contain any kind of functional specifications at present. The lack of geographic meaning stored against the cartographic features that Ordnance Survey holds is an obstacle to changing its production systems in view of challenges evolving around data interoperability, user focus and system flexibility (Regnauld, 2006). The overall aim of the thesis is to

conceive and build a model of functional information that can be used to process a topographic database to derive a richer thematic content. The localisation of extents of land use information, for example residential areas, can exemplify typical hidden information within the spatial arrangement of features. Indeed, such information is available in various forms. For example, local authorities or the Office for National Statistics provide freely available data, either statistical or spatial, on a range of topics including land uses (ODPM, 2005). Then there are commercial products such as the GeoInformation Group's Cities Revealed Image to Information database that consists of land use mapping, building height and building classification (GeoInformation Group, 2004). Even Ordnance Survey caught on to the need of a national land use database that offers the necessary content, coverage, level of detail, accuracy, consistency and diversity of data formats (see chapter 2). However, previous attempts have been hindered by limited and unsuitable data structures, and suffered from time consuming solutions that require a lot of human interaction, for example on-site surveys or manual interpretations of aerial photography. With the assumption that land cover configurations imply land use information, there is opportunity for exploiting automated methods. The importance here lies with the development of a method for automated enrichment of data concepts rather than its type of derived concepts.

The first objective is to define and specify the problem of enriching spatial data. The thesis investigates the potential meaning associated with the spatial arrangement of features represented in a topographic database. An object on its own does not mean much, but seen in its context to other objects it reveals a lot more information. This information shall be exposed with the help of knowledge representation. We need to ask if this has been possible before, what evidence there is that spatial configurations imply land use, and how this information can be determined through automated analysis. This includes the identification of functions that would add value to a topographic database, because its use is ultimately driven by what users want.

Second, this thesis analyses the process of map interpretation from a conceptual viewpoint to understand the problem and to specify a solution. The analysis takes into account spatial perception, cognition and categorisation to develop a suitable way for knowledge acquisition and elicitation. A questionnaire survey serves as a tool to impose a more rigorous structure on the process of eliciting cartographic knowledge. Together

with other human experiential accounts of cognitive geography, this investigation will indicate what types of knowledge are required for the inference of functional information. This knowledge then needs to be encapsulated in a semantic model for processing a topographic database. The survey provides the background for estimating the performance of any practical application to automated interpretation of land use information. Further, it is identified how the specified solution is a new step and whether it is an isolated effort.

The third objective is to design a conceptual framework that allows expressing and modelling any functional concept. It is important that the model accounts for the different types of properties and objects needed to discriminate different classes of functional information. These will be defined in a hierarchy of different levels of abstraction with the help of an ontology. The question is how can we achieve a link between the conceptual model, grounded in the conceptualisation of human physical experience, and the underlying spatial data representation? Will knowledge representation alone suffice for a solution so that we can completely abstain from using any black box, procedural knowledge? With that in mind, I propose representations for the new types of information, and ways to generate them automatically from the data through semantic reasoning.

The fourth objective is to implement and verify the solution by modelling the high-level concept 'residential area' within the conceptual framework. A concrete application of the axiomatisation of the required knowledge will be developed with the help of first-order predicate logic that provides the formal definition for the inference problem. The aim is to illustrate how the model pulls out implicit information from Ordnance Survey's topographic database using Protégé. Protégé is a free open-source Java Ontology Editor that provides an extensible architecture for the creation of customized knowledge-based applications (Gennari *et al.*, 2003). From this, we learn more about this particular contribution in terms of its limitations, both conceptual and technical, its purposes and potential uses.

Ontology has been among the most thriving themes in geographic information science over the last couple of years. The popularity of the research topic is highlighted by the share of papers published in competitive GIScience research outlets that deal with

research in geo-ontologies. According to the UCGIS research agenda of 2006, predominant issues are (a) the design of geo-ontologies from upper-level ontologies to domain-specific application ontologies, (b) methods for exploiting geo-ontologies for spatial querying, in particular spatial similarity searches, (c) the role of geo-ontologies within the geospatial semantic web, (d) and methods to manipulate geo-ontologies such as aligning and fusing them. Ontology is not a purely esoteric topic, but one of immediate interest to industry. For example, Ordnance Survey invests much research in this field in terms of enabling data integration by describing its data to their users' understanding and providing a framework for specifying product content. Ontology is also exploited to formalise the language of cartographic elements and to implement formalised mechanisms to enrich cartographic representations for on-demand mapping.

The value of using ontological descriptions lies in the ability to formalise knowledge and representations. The thesis addresses the intuitive understanding of relationships between geographic features and their meaning. Despite the difficulty to articulate and therefore to formalise this relationship, the use of an expressive representation language offers flexibility and explicit modelling of the problem. The conceptual framework is designed to dissociate itself from the underlying spatial data, and therefore can be applied to different datasets. The model not only extends easily by incorporating new concepts, hence facilitating data interoperability, but its derived concepts enrich the source data for further exploitation. To summarise, the proposed method contributes conceptually in the following ways:

1. The thesis studies and defines special issues that arise in processing a topographic database for the inference of higher-level information. It describes a framework centred on description logics for concept-based instance retrieval. The novelty lies in the method's conceptual and technical abstractions that facilitate the interpretation of functional information within topographic data.

2. The thesis defines the conceptual, or semantic, generalisation of the concept residential area. It describes how higher-order concepts are deduced from simpler ones by linking rich knowledge to spatial data in support of reasoning and inference. The specialised knowledge and reasoning skills of persons are emulated with a knowledge-based system to support automated map interpretation and information extraction. The thesis exploits existing work from

the fields of computer vision (e.g. Neumann and Möller, 2008), but contributes with its transparent, derived data concepts for cartographic representation.

3. The conceptual framework is dissociated from the data, but allows for a direct link. This ensures the framework will not be affected when changes occur to the source data. The described concepts within the framework serve as hooks to the data and can be easily related to concepts from other domains, hence contributing towards data interoperability. This is particularly relevant to research at national mapping agencies, which looks at ways to bind together different components of an on-demand mapping system facilitating the integration with third party data.

From a technical viewpoint, the thesis contributes by illustrating the solution with an off-the-shelf software application – the standard ontology editor and knowledge acquisition system Protégé 4 Alpha. More specifically, the contribution is an implementation that follows the structure recommended in the conceptual framework. This includes the following aspects:

1. A description logic application study is valuable per se due to the intellectual complexity of the field. The study formalises higher-level concepts enabling logical reasoning. It provides alternative solutions for overcoming the spatial representation problem within description logics, by linking concept definitions to spatial analysis methods.

2. The thesis defines a systematic approach of converting measurable spatial database properties into high-level semantic information. An expert system formalises data hierarchies for the interpretation and identification of regions with associated functions at different levels of abstraction. The conceptual definition of classes allows for the flexible extraction of instances, and provides an illustrative and explicit modelling of information.

3. The thesis demonstrates how aggregate concepts are inferred from lower level specifications using semantic rules and definitions. It implements these for the transformation of topographic data at different levels of abstraction.

Considering that GIS represents an abstraction of the real world in digital form, there are critical issues regarding the inclusion or exclusion of different forms of knowledge. Ontology, in its philosophical meaning, refers to the theory of existence. It asks the

fundamental question what exists, what are accepted facts and what can be known. In the context of GIS, ontological issues refer to what GIS researchers believe exist and how to represent this existence inside a computer. From an epistemological viewpoint, how space or place is defined inside a GIS affects not only what we can know, but how we know it (Sui, 1996). The world has been represented in various systems of GIS and geographic information that have evolved and been fashioned over time. However, GIS is a Cartesian model of space that excludes certain forms of representations. There is a strong argument that the focus of GIS needs to shift from representation and analysis of the form of the Earth's surface to a much stronger concern for the processes that define its dynamics (Goodchild, 2004). This means we need to distinguish between representation, a description of a given geographical location in space, and interpretation, a description of the displayed scene or mapping. On an ontological level, interpretation relates to different worldviews that give rise to a variety of visualisations and representations. We need to focus on the interpretive value of geographic information.

1.2 A guiding example

"He who searches for methods without having a particular problem in mind will most probably search in vain."

–David Hilbert

Urban areas encapsulate a wide differentiation between social, functional and morphological characteristics. From different classes and social groups, and different types of human activities and land uses, to the varying physical and spatial qualities – the relation between these factoring processes is evident. No matter how cities form, their spatial patterns are a reflection of physical, ecological and socio-economic processes within their boundaries and beyond (Luck and Wu, 2002). Therefore, the central object of theoretical thinking should be the physical and spatial form of our landscape, which yields important information in the form of indicators of attributes such as population density and composition, environmental impact, historical development and a range of cultural and symbolic dimensions (Pesari and Bianchin, 2000). To demonstrate this intricate relationship between spatial and functional patterns, figure 1 gives a concrete example from remote sensing and cartographic mapping. Both

the aerial photograph and the topographic map depict the same geographic location in an urban area. The objective is to derive a description of the scene based on our generic knowledge about the world and its types of objects and their characteristics. These traits can be identified as reoccurring patterns and are used to derive the function of the represented entities. The question to you, as the reader is, if you can identify the types of land uses depicted in the scene. The aerial photograph will yield familiar information in terms of texture, colour and representation, whereas the map is an accurate, but more abstract representation of the same features. Whether one is acquainted with mapping or not, by searching for familiar spatial configurations, it should not be too difficult to get an idea of the kinds of land uses present in the scene.
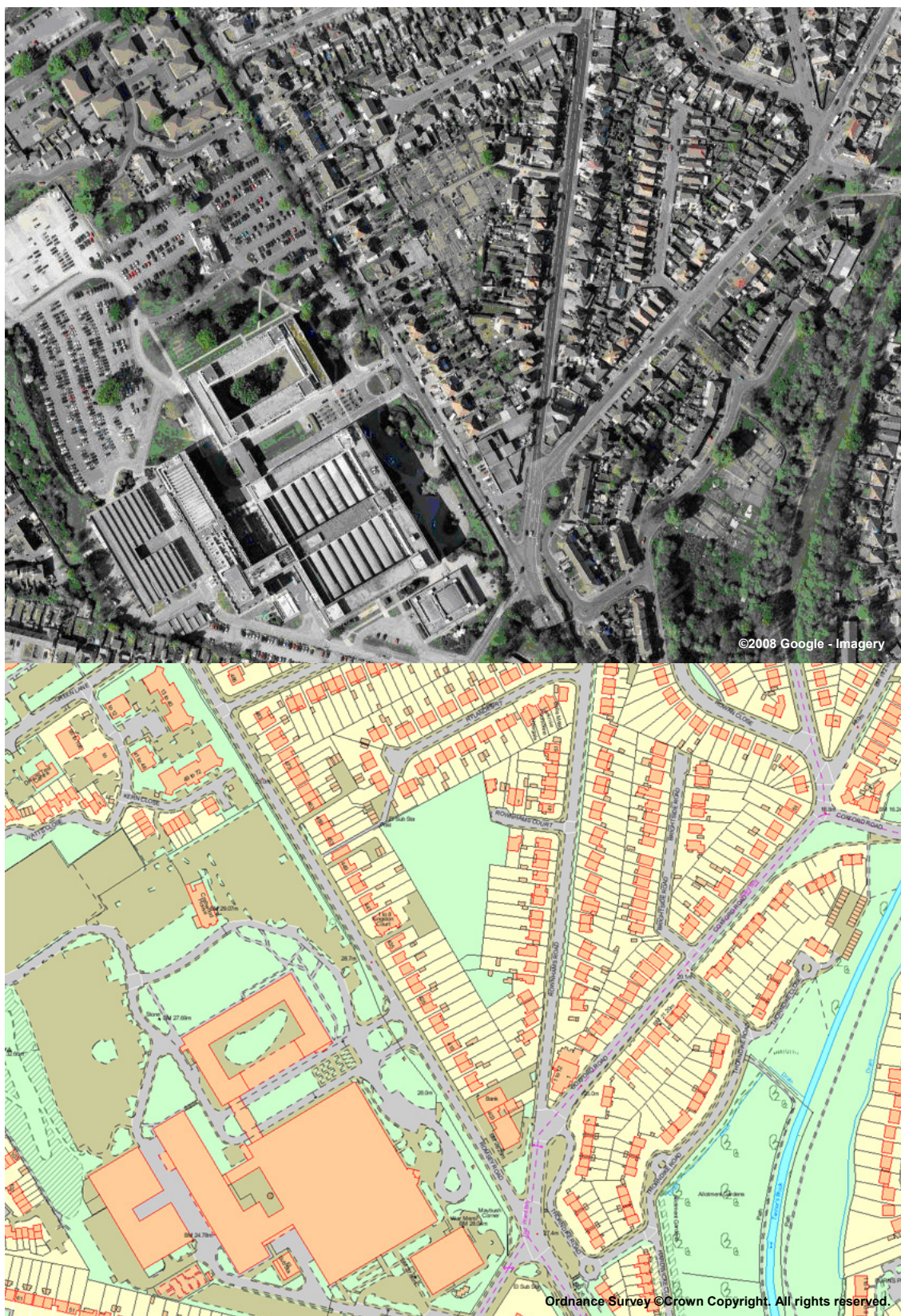
Figure 1 What land uses do you interpret in this scene?

The combination and interaction of simple patterns – whether presented in a continuous model of pixels, i.e., the image, or in an object-oriented view of vector data, i.e., the database – can lead to higher-ranking patterns or explicit new information that are hidden in the data (Heinzle *et al.*, 2005). Image interpretation is a long established discipline, especially in computer vision, artificial intelligence and remote sensing. It has produced many standard applications to recognise structures in pictures (e.g. Haralick and Shapiro, 1992). The human ability to account for context and to interpret information has been the inspiration for such undertakings. From the example above, it is clear that we can establish a direct relationship between socio-economic information and the spatial distribution of land cover parcels. You may have guessed correctly, the scene in figure 1 includes land uses such as residential area, including detached, semi-detached and terraced houses, green spaces such as a park to the right hand side, and some sort of industrial complex with large open-spaces such as car parks and grass areas. What you probably have not been able to infer is that, more specifically, the scene shows Ordnance Survey's headquarters in Southampton on Romsey Road.

The usefulness and necessity of interpreting vector data is still undervalued (Heinzle *et al.*, 2005). Land use data generally use parcel databases as a spatial framework (Wiegand *et al.*, 2002). The topographic polygons in Ordnance Survey MasterMap provide such a basis for defining and classifying individual units of land use. For instance, the text names and descriptions associated with real-world objects in OS MasterMap provide already a rich source of information that can be used to help derive land use descriptions (Wyatt, 2004). A land use theme would allow users to extract a set of features that are members of the theme definition allowing the selection of whole land use areas. Still, the main issue is that geographic extents of land use activities do not form a neat two-dimensional mosaic of polygons. In reality, a land use activity may encompass a complex of land and building polygons that are interrupted by roads and other infrastructure, such as an airport. Alternatively, there may be several different uses on different floors of a single building, such as a ground floor retail unit with residences above. The segmentation of continuous data representations, by dissecting heterogeneity between pixel values, equally suffers from the non-homogeneous distribution of land use patterns. The outcome is often low interpretation accuracy (Mesev, 2003; Hansen, 2003). As our environment becomes increasingly modified resulting in an ever more fragmented landscape of more and smaller patches, quantitative as well as qualitative

spatial analysis methods are needed (Luck and Wu, 2003). Based on economical principles and conceptualisations by humans, typical patterns can emerge in the spatial arrangement of land cover parcels. We can use these patterns to find regularities and to combine different rules or occurrences of special structures to determine general land use types. This raises a number of questions.

Research Questions

1. *What can spatial context and its configuration tell us about the functioning of its features?* – This question is important to identify if additional information can be inferred from spatial representations. However, the study of architectural form and spatial pattern within societies is far from simple, due to the bewildering distribution of similarities and differences that cause variations among spatial form of settlement structures, and hence prohibit their categorisation and generalisation. By answering this question, we will learn about the theoretical implications and existing practical applications that derive land use information from spatial data. We will identify why it is important to enrich spatial data in such way, and discern existing options, their limitations and how we can improve upon those, in chapter 2.

2. *What can we learn from our own abilities to interpret land use information from topographic maps? What kind of knowledge and reasoning processes are required?* – The interpretation of information is a knowledge intensive task. By finding persons with a reasoning skill that is important to the problem at hand, talking to them to determine what specialised knowledge they have and how they reason, we can embody that knowledge and reasoning in a program. The relevant material will be provided by a questionnaire survey aiming to derive consistent cognitive information from human experience of geographical space in chapter 3. This kind of investigation will elicit the key processes and factors involved in deriving information, and we will learn how people conceptualise this type of information. This is crucial for knowing what needs to be represented in a model for processing this kind of knowledge.

3. *How can people's knowledge be captured and transformed into machine-readable format?* – Implicit information exists on the level of the relationships between geographical features, their extent, density, uniqueness and more. This knowledge often is well known by humans with their cognitive abilities, but has

to be made explicit for the computer. Whereas many spatial inferences may appear trivial to us, they are extremely difficult to formalise so that they could be implemented on a computer system. Translating those key processes and factors into a machine-comprehensible form requires knowledge representation formalisms. Ontology is a formal conceptualisation of the world. It specifies a vocabulary that uses a set of assumptions regarding the intended meaning of its words. These concepts can be formalised in machine interpretable way through description logic languages that provide formal foundations and reasoning support for expressing axioms and constraints on the concepts in the ontology. In chapter 4, we will learn how this translation can be achieved and how spatial data can benefit from such an approach.

4. *How can we bridge the gap between knowledge, i.e. conceptualisation, and geographic data, i.e. representation?* – Once the knowledge is available in machine interpretable form, it needs to be linked to the topographic database so it can be used for processing. This is not an easy task considering the principle of databases has been storage of non-redundant data to avoid potential inconsistencies. In addition, there are technical issues in regards to linking ontologies directly with a spatial database. However, the concepts and methods people use to infer information about geographic space become increasingly important for the interaction between users and computerised GIS. There is a big gap between what a human user wants to do with a GIS, and the spatial concepts offered by the GIS. Although formalised spatial data models have been extensively discussed in the context of databases and geographic information, there are no models for a comprehensive treatment of different kinds of spatial concepts and their combinations that are cognitively sound and plausible (Egenhofer and Mark, 1995). Therefore, from studying human spatial reasoning, we can deduce a conceptual hierarchy of different levels of representations that tie higher-level knowledge to the geographic data. In chapter 5, we will learn about the mapping between knowledge and data using the example of the high-level concept residential area.

5. *How can geographic space be modelled in terms of its context and arrangement?* – Not only do people employ several different concepts when thinking about geographic space, but spatial representations have several levels of granularity, i.e., scales. Reasoning in geographic space must typically deal

with incomplete information that requires one to intelligently compensate for missing information and to apply default rules based on common-sense reasoning. To capture any form of conceptualisation requires therefore sufficient representational power of a given representation language to model geographic space in all its complexity. Especially land use is a difficult concept to model because of its semantic and spatial ambiguities. Nevertheless, in chapter 6 we will learn how a representation language deals with spatial relations by interpreting concepts as sets of individuals, and roles, i.e., relations, as sets of pairs of individuals.

6. *What type of functional information can be derived from topographic data alone?* – This question is important in two ways. First, we need to identify which land uses can potentially be derived from their spatial configuration through a visual analysis based on the interpretative capabilities of people. Secondly, this knowledge is then represented in a conceptual model to computationally process topographic data. Chapter 7 discusses how implicit information is practically inferred from Ordnance Survey's OS MasterMap topography layer, and analyses how successful the approach is.

Answers to the above questions will aid to solve the overall research problem of enriching spatial data semantically through the exposure of new, previously implicit information. It is already becoming apparent that the thesis is stepping on *terrae incognitae*. It falls in between many, and yet squarely within none, of the social sciences. Perhaps the closest is cognitive science, which is itself a combination of other sciences like psychology, philosophy, linguistics, computer science, neuroscience and anthropology (Miller, 2003). In any case, more questions will rise from the thesis while we substantially develop and explore how categories of functional information and their levels of granularity can be inferred and represented within topographic data.

Summary of the thesis argument

1. Urban studies argue that space creates a special relation between function and social meaning, thereby relating spatial configuration to social structure. Spatial representations, such as topographic maps, thus implicitly store functional information through the spatial arrangement of its features.

2. The tradition of constructive spatial representation fails to match common human representations of the spatial world. Efforts to construct new spatial representations

that successfully match human cognitive perceptions, rest on the ability to relate people's subjective interpretations with an artificial machine understanding of land use. The semantic enrichment of spatial data, therefore, improves accessibility, definition and suitability to a wider range of applications.

3. Human experiential accounts of cognitive geography exist, and are available for machine collection, aggregation and semantic interpretation through commonsense understanding. Together with a specific investigation of how people interpret land uses from topographic maps, this indicates the required reasoning processes and knowledge for enriching topographic data.

4. Ontology provides the foundation for knowledge bases. It implements through logical theory as a conceptual system for data integration and information retrieval. With its ability to structure and organise knowledge, to provide communication between humans and machines, and to reason by inference, we can use ontology as the mediating instance between the represented world's reality, i.e., spatial data, and the information that is required according to human understanding.

5. A conceptual model of residential area, grounded in the concept of affordance and the interpretation of generalised human experiential accounts, constructs new representations of space. Its hierarchy of concepts captures the semantic distinctions necessary for generating land use information from topographic data.

6. Representational formalisms with appropriate expressiveness for capturing the necessary facts, properties and constraints, translate knowledge into a formal and machine manipulable model. Due to the language's logic based semantics, a machine can reason about the asserted knowledge and infer higher-level, initially implicit information.

7. The conceptual model is implemented in OWL-DL with Protégé 4 Alpha to infer different types of dwellings and their extents from Ordnance Survey's topographic database. The link between data model and conceptual model is achieved by having the application classify the data's knowledge (i.e., facts) in terms of the general semantic categories that the conceptual model (i.e., ontology) provides.

8. This approach is shown to offer powerful means for enriching spatial data, enabling reasoning over its asserted facts, integrating with other data sources, and providing concepts that follow human intuition and understanding. With the ontology's expressive power, there is indeed potential to derive land use information from topographic data alone.

1.3 Thesis structure

In the following chapter, known approaches to deriving land use information from spatial data are reviewed in the literature, and the problem that this thesis addresses is further defined. Chapter 3 then analyses the problem from a human conceptual point of view, reviewing implications from spatial cognition and categorisation, to understand the types of knowledge required for inferring hidden information from topographic maps. Thereupon, a methodology is proposed to automate this process in chapter 4, which introduces ontology and evaluates how this approach benefits spatial data in terms of data enrichment and integration. In chapter 5, a conceptual model is deduced that defines a theoretical hierarchy of concepts and rules that allow inference of higher-level semantics from a topographic database. Because the functional geography is too complex to be modelled as a whole within the scope of this thesis, the model and implementation thereafter focuses on the concept residential area. Both conceptually and spatially, residential area is a straightforward concept with typical, easily discernable patterns that are ideal for illustrating the proposed solution. Chapter 6 then addresses how this conceptual model can be formalised with the help of representational languages such as description logics. The use of long-established theories such as logic provides a sound foundation for translating the acquired knowledge into machine-readable format with which topographic data can be processed. The inference of implicit information is achieved through concept-based instance retrieval, one of the basic reasoning tasks that a representational language can easily perform. Chapter 7 evaluates the proposed approach with the freely available ontology editor Protégé. The conceptual model is implemented with the Web Ontology Language (OWL) and is applied to OS MasterMap topography layer. With the facts from the topographic database asserted in the knowledge base, topographic instances can then be classified according to the concept definitions given in the model. The results are analysed and the approach is assessed in terms of its strengths and weaknesses. Chapter 8 closes the thesis with the recapitulation of its contributions as well as an outlook for possible future investigations.

Chapter 2

Definitive, Intelligent and Accessible Data

"...GP surgeries, large food shops, primary schools and post offices. These are all key services that are important for peoples' day to day life. In this sense wherever you live, having to travel a long distance to such places can be described as an access deprivation."

–Office of the Deputy Prime Minister (2004, p.31)

The growth of cities is dependent on responsible and sustainable use as well as efficient and ethical planning to ensure that our land can cope with the increasing populations in future. The growing out flow from cities to the edge of town and the countryside, the dispersion of key services, and neighbourhood deprivation in general, are challenging tasks for the government of today. Geographic information forms an essential part not only for studying urban form and the spatial distribution of poverty (Vaughan *et al.*, 2005), but to combat exclusion and update the indices of deprivation, such as the 'Barriers to Housing and Services Domain' (ODPM, 2004). Relevant, up-to-date, consistent, detailed and geo-referenced information are required to target solutions more effectively on the most deprived neighbourhoods. For example, Ordnance Survey (Harding, 2003) acknowledges that new concepts including land use, social resources, and neighbourhood, as well as new geographical entities, for instance GP surgeries, pharmacies, food shops, etc., are required to meet the needs of social policy makers.

This illustration highlights the importance of definitive, intelligent and accessible data. Access is linked to skill and knowledge, and therefore addresses both the lack of interpretive skills by individuals as well as data sources with insufficient thematic content. Most GIS, for example, require extensive training, not only to familiarise the users with terminology of system designers, but also to educate them in formalisations used to represent geographic data and to derive geographic information (Egenhofer and Mark, 1995; Haklay and Zafiri, 2008). The question is when we access data, what are the capabilities and possibilities for empowerment? Data alone only present symbolic representations of realities. It forms the most elementary level where data contexts are merely considered as formal set structures without any content (Stenmark, 2001).

Information is a collection of related data that have been processed into a format that is understandable by its intended audience and presented in a form that is suitable for human interpretation. It is information, 'data endowed with relevance and purpose' (Jerome, 2003), that makes data useful and ultimately leads to knowledge, the ability to utilise the information effectively. Knowledge is what enriches the use of data the most by assigning meaning. In technical terms, knowledge comprises a body of organised information in a context that guides action based upon insights and experiences.

In the context of the thesis, spatial data refers to geographic information, and semantics refers to the processing of knowledge using declarative languages to define what entities mean with respect to their roles in a given system. There is a lot of potential knowledge stored within spatial datasets, both explicitly in terms of collected geometrical features and associated non-spatial attributes, and implicitly in terms of topological information, typical structures between features and relations of their attributes (Heinzle *et al.*, 2005). Knowledge forms the crucial element to expose new information by modelling the structures we want to recognise in the data (Lüscher, 2007). Consequently, by defining the semantics of functional information in relation to their spatial configuration and other clues stored in the data, we can augment existing databases with new knowledge and thus increase their value.

To achieve data enrichment, one needs to understand the theoretical and practical implications of what can be inferred from existing spatial data. This chapter investigates the relation between the spatial arrangement of land cover parcels, i.e., topographic features, and functional information. From the conceptual viewpoint, this chapter reviews studies from urban morphology and structural anthropology to learn more about the form and spatial pattern within societies. From an application viewpoint, the chapter discusses existing methods from the fields of knowledge discovery, generalisation and remote sensing to identify if land use information can be derived from an analysis of its land cover distribution, and how we can improve on this. Section 2.2 defines inference as a configuration problem, and what implications this holds for defining a solution. Section 2.3 relates the thesis to research at Ordnance Survey (OS), the national mapping agency of Great Britain. With the help of previous research carried out at OS, this thesis identifies which types of functional information would add value to a topographic database.

2.1 The functional landscape: From land cover to land use

The understanding of the messy irregularity that characterises the patterning of the real world is challenging to any model. Urban areas are considered as organisational entities with all kinds of intricate interrelations between their elements. We are interested in the interrelationships between land cover and land use. Land cover refers to the surface cover on the ground, whether vegetation, water, build-up areas or other. Land use refers to the purpose the land serves such as agriculture, industry or recreation (Hansen, 2003). The starting point is the assumption that land uses have specific organisational patterns.

Theories

The notion that space creates a special relation between function and social meaning has been long established in urban geography. Cities are socio-economic systems, and urban analysis therefore is at least as much about human activity patterns as it is about the built environment (Longley and Mesev, 2000a). The study of the social patterning of urban areas developed in the Nineteenth Century around typologies of the structure and configuration of society (Pesaresi and Bianchin, 2000). Morphological analysis was intended to be an explanatory tool, a means of relating spatial form to generating process (Longley and Mesev, 2000a). Many classic theories have blossomed from urban analysis with an attempt to study the morphology and evolution of cities, including concentric zone theory, sector theory, multiple nuclei theory, as well as recent theories such as catastrophe theory, chaos theory, dissipative structure theory, fractals, cellular automata, and self-organisation (Luck and Wu, 2002). Judgement about what constitutes 'good' urban theory is relative to what is known already. It is evident that better measures of urban phenomena based upon a better digital data infrastructure can lead to better description and thence to better theory (Longley and Mesev, 2000b). This has evolved from dealing with problems of simplicity, via the ability to deal with problems of disorganised complexity, to the analysis of cities as organised complexity problems (Pol, 2002). For example, the urban modelling tradition of the 1970s was neither able to come to terms with the countless forms of human agency, nor the serrated irregularity of urban morphology that arises out of urban growth dynamics in the real world (Longley and Harris, 1999). This meant traditional models became increasingly irrelevant to understanding city systems, and with that the quest to relate form to function, patterning to social process, was largely abandoned. Since that time, urban geography has arguably been overwhelmed by the task of representing the statics

and dynamics of spatial structure. The more recent history of urban modelling has been of more successful prediction and forecasting, achieved through the analysis of relationships between a considerable and manageable number of variables that target a specific problem (Longley and Mesev, 2000b; Pol, 2002). Only when the complicated social, economic and cultural interrelations within urban regions are understood and analysed, can we soundly tackle urban problems. Unfortunately, our ability to develop understanding of physical and socio-economic distributions through urban modelling remains limited by the quality and scope of data available (Longley and Mesev, 2000b).

Complex theories have developed from the obscurity of the problem. They are a clear indicator of the challenging prospect to study architectural form and spatial pattern within societies, due to the bewildering distribution of similarities and differences that cause variations among spatial form of settlement structures, and hence prohibit their categorisation and generalisation. As researchers describe their acceptable designs as fit between urban form and its context, many disputes have risen over the description of the urban structure. For example, Alexander (1965) sees the city as much more than a simple system of units, or sets of elements, that neatly divides functions from each other. Instead, the city portrays the overlapping nature of activities with all kinds of intricate interrelations between their elements that indicate a living system. Therefore, he greatly opposes the hierarchism of the city, and questions the approach of reducing cities to hierarchical classifications and graph-theoretic sub-divisions of urban elements, because the structural simplicity and lack of interconnection between units within a tree confines, and in fact cripples our conceptions of the city. According to Alexander, land uses in cities are composed of overlapping areas whose order is more lattice- than tree like. He reasons that the cause for an adoption of the tree structure by so many is the limited capacity of the mind to form intuitively accessible structures that cannot encompass the complexity of the semi-lattice in a convenient form and single mental act. Further, the tree conception leads to compartmentalisation and the dissociation of internal elements and, hence, implies separation and destruction. Although Alexander does not give much credit to the human conceptual capacity, he is right in one point. Many idealised concepts of urban configurations developed by city planners and developers, such as the Garden City (Ebenezer Howard), the Radiant City (Le Corbusier), and the Broadacre City (Frank Lloyd Wright) adopt an underlying tree-like structure to their functioning. In reality, however, we are dealing with diversified cities,

where activities can be of primary or secondary use, and some are even combined (Pol, 2002). This reflects the obscurity of the problem, that function can be perceived differently, i.e., for some an activity is primary while for others is secondary, and therefore emphasizes Alexander's argument of overlapping activities and that cities should be treated as problems of organised complexity. However, the quest to understand function has involved ideas of hierarchies and networks, and the search for functions that are consistent with the shape of cities and their evolution. Many models, such as spatial discrete choice models, spatial interaction-entropy models, and standard multivariate cluster-type techniques therefore employ the notion of hierarchical urban structures (Batty and Longley, 1994). In fact, difficulties in obtaining objective and consistent definition of categories of urban land uses have been identified, and that their level of complexity threatens to destroy the most sustained attempt to classify their geometry. This shows that despite the acknowledgment of the problem, still often a simplistic approach and treatment is taken to analyse and understand the urban fabric.

Form and function of space are dependent on one another. As the phrase 'form follows function' states, "*various processes which contain the forces which determine form have specific functions and a study of form from the static viewpoint, from one snapshot in time for example, is often rooted in the quest to understand functions*" (Batty and Longley, 1994, p.42). Space is a shape, and function is what we do in it (Hillier, 1996). This relationship goes both ways. The environment and what it affords, that is, what objects or things offer one to do with them (Jordan *et al*., 1998), guide both the perception and action of people. This not only endows the layout of the environment, or its physicality, but a complex, information-rich, ever-changing environment, which is furnished with cultural objects and characterised by social interactions that it affords. The use we make of an environment thus is depended on what it allows one to do as well as its deliberate purpose that supports some type of function, which may or may not be realised through its specific use or role. Therefore, if affordances influence the functioning of our environment, then this functional environment will have an effect on the human behaviour, which, shaped through the sensory inputs and intrinsic information manipulation, makes use or takes place on that environment. From an urban perspective, this results in two problems. The multifunctionality of cities means that every aspect of the spatial and physical configuration works in many different ways, influenced climatically, economically, socially or aesthetically, with form only changing

slowly while function changes rapidly. Secondly, cities are made up of parts with a strong sense of local place, which clouds the morphological distinction between one part and another. Hence, two problems of description are needed and must be solved according to Hillier and Hanson (1984). That of society, which is to be described in terms of its intrinsic spatiality, and that of space, which is to be described in terms of its intrinsic sociality, in order to define the social logic of space and the spatial logic of society, respectively. Hillier (1996) argues that these two issues are really the same problem, because the fundamental correlate of the spatial configuration is movement, where movement largely dictates the configuring of space in the city and vice versa. Movement forms an integral part of the physical environment and human behaviour. Because urban space is a place of potential meetings and interactions of its inhabitants (Wlodarczyk, 2005), the built environment is not merely a material backdrop to individual and social behaviour. Both cultural and social ideas are transmitted through configuration, that is, the raw materials of space and form are given social meaning. It is evident that Hillier assumes that people and societies deploy themselves in space, and that these deployments are capable, under certain conditions, of adopting certain patterns. Functions are a result from these deployments. They are embedded as relationships between spatial configurations as a whole, and one will find that it is common that different functions are 'spatialised' in different ways. Hillier (1996) claims that the analysis of regularities in the relation between spatial configurations, defined as a set of interdependent relations, and the observation of the functioning of the environment allows the discovery of the distribution of land uses, such as retail and residences in an urban context. However, the difficulty in analysing settlement structures lies in the lack of well-defined spaces with well-defined links from one to another, because of its continuous structure of open space, which is not easily decomposable into elements for the purpose of analysis.

Applied methods

Hillier and Hanson (1984) introduced the configurational theory with the space syntax model, which describes society by associating social theory of production with the use of space patterns. The resulting urban space is a reduction to a complex of lines, or axial lines, and convex spaces. Their theory has been taken further by numerous authors (e.g. Cutini *et al*., 2004; Perdikogianni, 2003; and Kasemsook, 2003; Béra and Claramunt, 2004) to study practical examples of spatial and functional pattern, and to perform

configurational analysis. Graph-theoretic representations became the norm for articulating urban space as a pattern of identifiable urban elements such as locations or areas whose relationships to one another are often associated with linear transport routes such as streets within cities (Batty, 2004). However, as in any theory, there is criticism. For example, space syntax techniques hardly bear any sign of the morphology of the urban spaces that gave rise to the lines in the first place, hence losing trail of the wide open spaces such as squares (Cutini, 2003). Furthermore, space syntax does not directly take into account land use factors, but instead keeps these separate to investigate the impact of both configuration and movement on land uses. According to Hillier and Penn (2004) space syntax expects land use to be a dependent variable, because if spatial configuration influences movement it can be expected to influence land use patterns with respect to their demands for being close to or avoiding movement. Nevertheless, the significance of Hillier's research to this thesis is that according to configurational theory, the spatial structure of a settlement, that is the way its streets and squares are disposed and mutually related, is the actual key for the comprehension of urban phenomena, both material and immaterial (Cutini *et al.*, 2004).

Considering the intrinsic relation between spatial form and function, it becomes clear that any kind of quantification of spatial heterogeneity in spatial data requires a way to describe and represent variability in space and time (Gustafson, 1998). A number of approaches exist from disciplines such as data mining, where implicit information is exposed by using spatial rules to extract regularities within the data (Koperski and Han, 1995; Lu *et al.*, 1993). Automated map generalisation, for example, addresses problems of pattern recognition and structure modelling for preserving structural properties of a set of objects when these are generalised at different scales (Mackaness and Edwards, 2002; Jiang and Claramunt, 2004; Zhang, 2004). Many of the recent advances in the computational recognition and analysis of spatial patterns present in geographically referenced digital data sets have used approaches from the broad field of pattern recognition (Barr and Barnsley, 1998; Chou, 1995). Although Hillier (1996) resists the word pattern, because it implies more regularity than one will find in most spatial arrangements, the identification of such spatial arrangements indicates their use. For example, residential areas in many Western European towns and cities typically form a complex compilation of buildings (houses), roads and open space (gardens and parks). On an aerial or satellite image, the thematic and morphological properties of these

parcels (their land cover type, size and shape) together with the spatial and structural relations between them (e.g. adjacency, containment, distance and direction) convey information on the associated land use (Barnsley and Barr, 1997). What is required, then, is a set of techniques that exploits these sources of information in an automated or semi-automated manner to infer land use from the spatial pattern of land cover. Urban areas offer a particular challenging landscape, as they are organised spatially into irregularly shaped land parcels of buildings, roads and various types of intra-urban open space. However, a separability analysis of land use samples from remotely sensed imagery suggests that a quantifiable mapping exists between urban form, i.e., land cover, and urban function, i.e., land use (Barr *et al.,* 2004).

The two main types of data structures for representing real world scenarios are vector data and raster images, of which the latter has by far the more developed techniques for image interpretation and inference of new information based on the recognition of structures within pictures (e.g. Mather, 1999). The process of deriving thematic information from digital, remotely sensed images is commonly based around the use of per-pixel, statistical analysis or artificial neural network classification techniques (Barr and Barnsley, 1997). The general approach is to identify the dominant land-cover type associated with each pixel and then examine the spatial arrangement of these land-cover labels in multi-pixel regions of the image. For example, in a remotely sensed image of urban areas, residential land typically consists of a complex spatial assemblage of tarmac and concrete roads, slate and tile roofs, trees, grass, shrubs and bare soil, each of which exhibits a different detected spectral response (Barr and Barnsley 1997). Many categories of urban land use have a characteristic spatial pattern of spectrally distinct land cover types that enables their recognition in fine spatial resolution remotely sensed images (Barnsley *et al.*, 2001). However, such images are increasingly segmented into discrete, labelled regions closely related to the principal spatial entities in the corresponding scene. An object-oriented representation provides morphological information and allows us to quantify, interrogate and analyse the structural properties of the regions and the spatial relations between them. This information is crucial for an understanding of spatial configurations, that is, how characteristic patterns in a set of phenomena can be recognised by reference to abstract principles of arrangement or relationship.

Configuration is a set of relationships among things all of which interdepend in an overall structure of some kind (Hillier, 1996). Analysing configurations therefore requires a thorough understanding of what possible spatial relationships are among objects and how they can be determined. Max Egenhofer, for example, (1989; Egenhofer *et al.*, 1991; Egenhofer and Herring, 1991) defined binary topological relationships between n-dimensional spatial objects embedded in n-dimensional space, which allow the calculation of touching or overlapping objects. However, these relations are purely based upon topological properties – a taxonomy of all possible combinations of boundaries and interiors of two objects – and thus are independent of the existence of a distance function. Therefore, topological relations alone are not sufficient to provide a full description of a scene, and further relations have been defined in terms of distance and direction properties, especially within the field of qualitative spatial reasoning (Clementini *et al.*, 1997; Cohn and Hazarika, 2001). This means both spatial relations as well as other types of spatial characteristics, including unary object descriptors, ratio-type relations and/or attributes specifying the semantics of the spatial objects are required.

There are many techniques to establish patterns and infer higher-order meaning from spatial data, whether from images or vector data. This not only proves the general demand for improved information retrieval and exploitation of implicit information from existing data sources through automated reasoning procedures, but also that the scientific communities believe in a link between spatial form and higher-order meaning such as function. For instance, rule based aggregation uses a formalised set of rules for classifying the structural composition of objects. Bauer and Steinnocher (2001) use this technique to achieve the transition from the spatial distribution of land cover objects to land use entities. They treat residential area as a composite of large built areas adjacent to either medium grass or a medium tree area, whereas industrial areas comprise large built up and large open-space paved areas. The difficulty of this approach lies with the definition of such rules and the control strategy to infer new data from it, despite the development of concepts to integrate learning techniques for deriving the necessary knowledge. If, however, such rules are known or models of the situation are available, good results can be achieved.

As mentioned earlier, graph-theoretic approaches have proven popular for model-based interpretation derived from graph representations. Heinzle *et al.* (2005), for example, take such a graph-based approach to evaluate road network patterns within urban settlements and to determine the city centre. Barr and Barnsley (1997 and 1998) developed an extended relational attributed graph to infer second-order thematic information about a scene. This was developed further by de Almeida (2007) to infer higher-level information from unstructured datasets such as LiDAR data. Albeit existing techniques for assessing the similarity of graphs, for instance graph similarity measures and graph matching algorithms (Conroy Dalton and Kirsan, 2005), the problem arises with large data sets of greater structural complexity that will lead to computational inefficiency and greater uncertainty due to the much greater variability in the graph structure (Barnsley *et al.*, 2001). Similarly, the concept of gradual change, which originates from the gradual deformation of spatial objects until the spatial relation between them is changed, experiences the same problems (Egenhofer and Al-Taha, 1992; Bruns and Egenhofer, 1996).

Often these approaches are combined with clustering procedures. Anders *et al.* (1999), for example, apply graph-clustering techniques for the analysis of settlement structures by delineating homogeneous structures in a data set. The drawback is the requirement of prior information, such as the statistical distribution of the data or the number of clusters to detect. This means existing algorithms can break down if the choice of parameters is incorrect with respect to the data set being clustered, or the model did not capture the characteristics of the cluster. Furthermore, existing clustering methods tend to be closed and are not geared toward allowing the interaction needed to effectively support a human-led exploratory analysis (Guo *et al.*, 2002).

In remote sensing analysis, methods such as supervised classification, decision trees and moving window techniques suffer from the need for training datasets or a priori optimum size for the kernel, whose rectangular shape is often unsuitable for searching irregularly shaped land cover/ land use parcels (e.g. Barnsley *et al.*, 2001; McCauley and Geotz, 2004). Often interpretation accuracy is low, especially when using low-resolution imagery. The relationship between land use in urban areas and spectral responses recorded in satellite images is complex and thus precluding the use of traditional classification approaches (Hansen, 2003). With the aim to overcome these

limitations, combinations of traditional remote sensing techniques with other data and methods have been researched. For example, Harris and Longley (2000) integrate urban remote sensing with new, often commercial, sources of socio-economic data to augment satellite data and thus enhance models of the form and functioning of urban settlements. Mesev (2003) links Ordnance Survey's Address Point dataset with satellite imagery to study the spatial distribution of postal addresses based on density and arrangement to infer urban land use distributions of built environment, commercial and residential. Hvidberg (2001) and Hansen (2003) dismiss remote sensing altogether and instead employ the Danish Building and Dwelling Register database to create urban land use maps by overlaying a regular grid net and applying a fuzzy logic classification. Remote sensing techniques such as supervised classification have also been used in different contexts. Boffet (2000), for instance, applied the goal-directed classification for identifying homogeneous urban groups from topographic data. The classification is represented as a hierarchical tree with the final classes being predetermined according to the user's need. The difficulty here lies with choosing the adapted variables, measurements and threshold and with interpreting intermediate cases. In particular, the chosen thresholds have to be sufficiently sensitive to discriminate significant classes. To improve the classification of urban districts, Boffet employs the typology of properties of urban sub-systems that allow the analysis of shapes, structures and constructions, as proposed by Pinchemel (1995).

The kinds of higher-level information, which the exemplified techniques are trying to establish, closely link with the different conceptual levels of data description. The most fundamental and concrete level is raw data consisting of micro objects, such as a building, a road, or a community boundary. This is most useful at large-scale representations. At smaller scales, however, objects at a micro level of description are inadequate, as the desired objects become more abstract, called meso, such as a district, a block, or a town. The meso object is a relevant concept, as it describes the combination of micro objects based upon their collective representation of a geographic phenomenon (Boffet, 2000). Deriving the meso object from micro objects is not an easy task. The difficulty lies in the notion of geographical phenomena that require knowledge and the human capacity of interpretation in an interactive process. Interpretation generally is a knowledge intensive task, as we will further investigate in chapter 3. To improve previously developed techniques, which are mainly technology-driven, we

need to distinguish between the knowledgeable, i.e., the capability of machine learning, and recognisable, i.e., the human cognitive perception of patterns. Even Hillier *et al.* (1976, p.148) noted this key relationship as '*the relationship between the formal structure of what there is to be known (for example, the patterns of space organisation, patterns of social networks, and so on); and the formal mental structure by which these are known or recognised*'.

Computer vision embraces this relationship by automating and integrating a wide range of processes and representations used for visual perception. This requires the formulation of procedures and knowledge that encapsulate the content of the images. It is therefore widely acknowledged that research on information extraction must consider primarily the semantic aspects of the data. However, the complexity of the information stored in images makes this a non-trivial task. For example, a long-term research project at the University of British Columbia, called Mapsee, provided a first approach for a theory of image interpretation based on logic (Reiter and Mackworth, 1989; Mulder *et al.*, 1988). The Mapsee project experimented with visual knowledge representation from sketch maps of geographical regions. Matsuyama and Hwang (1985) developed also a logical foundation for their knowledge-based aerial image understanding system SIGMA. Approaches used in knowledge representation and modelling for machine vision have therefore been widely applied for image understanding of remotely sensed images (Sowmya and Trinder, 2000). But what about spatial databases containing vector data? The exposure of knowledge by means of knowledge representation, for example through ontologies, has been employed to improve retrieval of distributed geographic information (Lutz and Klien, 2006). Klien and Lutz (2005), for instance, use ontologies for the automated semantic annotation of geodata, whereas Hertog *et al.* (2000) applied a knowledge-based system for polygon classification. The semantic gap between low-level descriptors, such as images or other spatial data, and higher-level semantic concepts is achieved by using domain knowledge. This knowledge is often determined by the perception of an object, i.e. by the context, configuration, meaning and experience of the observer (Wertheimer, 1924). People can easily recognise spatial patterns; however, a translation of these cognitive processes into automated procedures is much of a different situation. Because knowledge-based systems integrate prior knowledge about the scene or several objects, they are strongly domain-dependent, and often do not separate knowledge from the procedures (Zlatoff *et al.*, 2004).

Looking back over the reviewed methods and their limitations, it is evident that there is a general paucity of formal models on the structural operators, their semantics and expected results, which are required to underpin the development of such techniques (Barnsley *et al.*, 2001). The existing lack of rigidity in definition and analysis in morphological studies arises in part because of the paucity of spatial analysis in providing measures of the distribution of physical and other spatial structures on the ground (Pesaresi and Bianchin, 2000). Furthermore, despite the acknowledgement of the problem of linking spatial configuration to the functioning of the environment by different disciplines, the focus seems to have been only on urban space and not on rural landscapes or non-urban scenarios. Non-urban landscapes and their pattern distribution are often only addressed in terms of landscape ecology to study links between ecological pattern and ecological function and process (Gustafson, 1998), or for strategic land-use allocation as it is the case for agricultural purposes (Carsjens and Van der Knaap, 2002). Therefore, it is important to designate the problem to any type of space, whether urban or rural, and to look at ways to retrieve functional information from various spatial arrangements. However, first we need to take a closer look at the underlying problem, that of configuration.

2.2 Treating inference as a configurational problem

Did you ever play with Lego^{TM} in your childhood? Imagine you have a set of Lego^{TM} blocks with which you want to build a small rowing boat model. This task translates into a set of requirements that specify general geometric properties of the boat and how its components must belong together. However, the type of Lego^{TM} blocks that are available to you and how they fit together, as well as the design of the boat itself, i.e., it should be symmetric and all part should be connected, constrain the design. The Lego^{TM} boat configuration therefore fully specifies the components, their shape and dimensions as well as their arrangement. The configuration problem consists of finding the optimal layout of the components, and involves some aggregation over the properties of the individual components. Given the standard dimensions of Lego^{TM} components, you will struggle to satisfy the requirements exactly. This is a typical phenomenon in configuration design (Wielinga and Schreiber, 1997).

Configuration problem solving has become an established field of Artificial Intelligence (AI) applications. It comprises the selection and instantiation, parameterisation, and composition of components out of a pre-defined set of types in such a way, that a given goal specification as well as a set of constraints characterising the configuration in general will be fulfilled (Klein *et al.*, 1994). Configuration is treated as a problem class with no indefinite goals, no unspecified constraints, with completely described objects, relations and constraints between them. In this sense, configuration is a well-structured phenomenon. The combinatorial nature of the problem requires problem-solving methods to constrain the search process of the relevant components. Knowledge about the configuration structure is therefore an essential precondition for the formulation of a solution, and provides a suitable platform for a formal description of the configuration problem. If the arrangement of components is an important element of the application problem, the solution can take the form of a hierarchical decomposition, where the top-level goal is decomposed into a number of alternative substructures (Wielinga and Schreiber, 1997). Generally, a knowledge-intensive approach is required to express knowledge – heuristic or otherwise – and to make logical inferences about the typical problem solving activities.

In geography, the problem is not as simple and sterile. Objects in a geographic space correspond to locations on the surface of the Earth, and complex geographic processes and structures can emerge from local interactions (Miller, 2004). Nevertheless, if built environments are considered as organised systems whose primary nature is configuration that expresses the social purpose for which the built environment is created (Hillier, 1996), then we should be able to apply methods of configuration problem solving. A first step would be to decompose the spatial configuration of land uses to understand the simultaneous effects of a whole complex of entities on each other through their pattern of relationships. By using this configuration knowledge, what logical inferences can we then make about a spatial environment? The answer lies in the contextual, structural and morphological properties of features represented in a spatial scene that together form a specific pattern or configuration. For example, consider a database in GIS that is composed of a large number of spatial objects that are spatially arranged and grouped into a few thematic layers, which usually cover the whole geographic space. You formulate a query to find a hospital in an urban area adjacent to a park and a highway. A query evaluation in a spatial database combines objects from

different thematic layers to construct desired answers based on these objects' spatial relations (Rodriguez and Jarur, 2005). This is a type of constraint satisfaction problem. Crucial to finding a solution is the identification of spatial relations. Indeed, as Tobler's first law of geography states, *'everything is related to everything else, but near things are more related than distant things'* (1970, p.236). Location has an intrinsic degree of uniqueness due to its situation relative to the rest of the spatial system. The emerging spatial heterogeneity reveals information on both the intensity and pattern of spatial associations (Miller, 2004). However, in GIS spatial relations are limited to expressions in the form of topological relations, distance relations, or orientation relations. The problem is that computer systems do not generally support context-based representations, and therefore spatial relations such as proximity, which are context dependent, fail to relate mental representations of space (Worboys, 1996). The work of Rodriguez and Godoy (2002) reflects this problem. They describe quantitatively spatial configurations and try to retrieve them automatically from spatial databases, however only with limited success. Any configurational modelling therefore needs to address two issues: The first issue is that of identifying regularities in the ways in which urban systems and their functions are put together spatially by identifying their genotypes of spatial form. An effective way to achieve this could be the use of pattern recognition methods (e.g. Hussain *et al.*, 2007). Secondly, these regularities need to be correlated with aspects of how humans conceptually interpret the functions observed in space. Take figure 2 for example, the human eye can easily interpret the spatial footprint of land uses with or without explicitly stating which land uses are depicted. This is because we have a common understanding of how the world works, and we are able to use implicit situational information, or context, to make sense of everyday situations (Dey and Abowd, 1999).
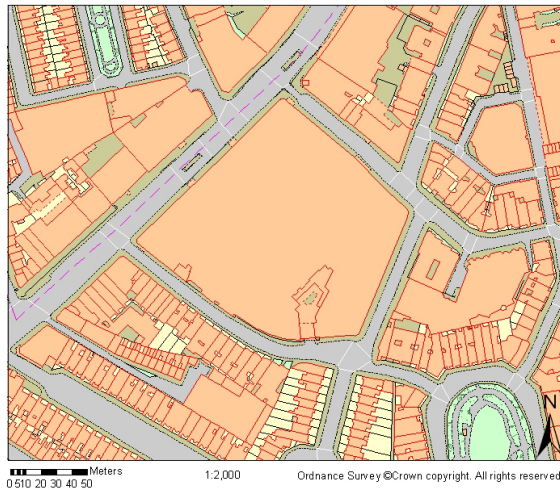
(a) Department store Harrods in Knightsbridge, London
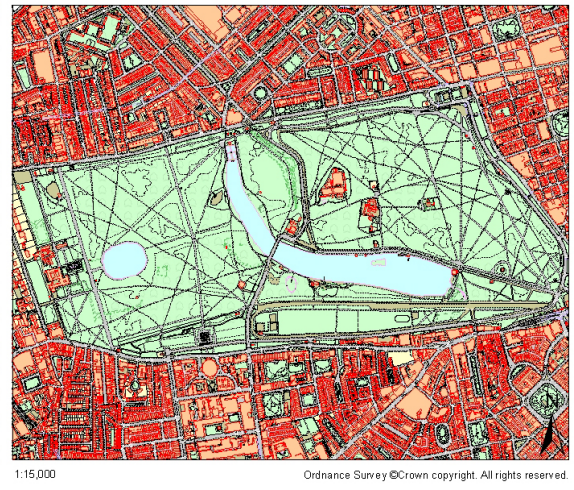
(b) Hyde Park in Knightsbridge and Belgravia
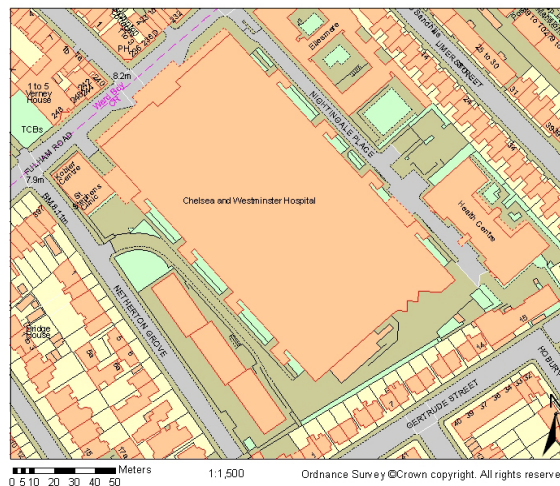
(c) Chelsea and Westminster Hospital, London
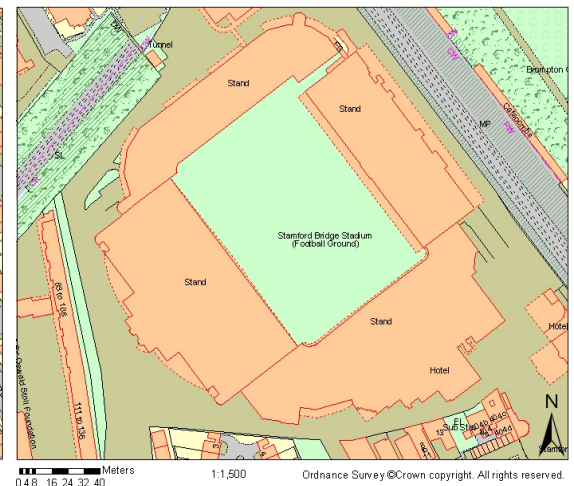
(d) Stamford Bridge Stadium, Chelsea Football Club
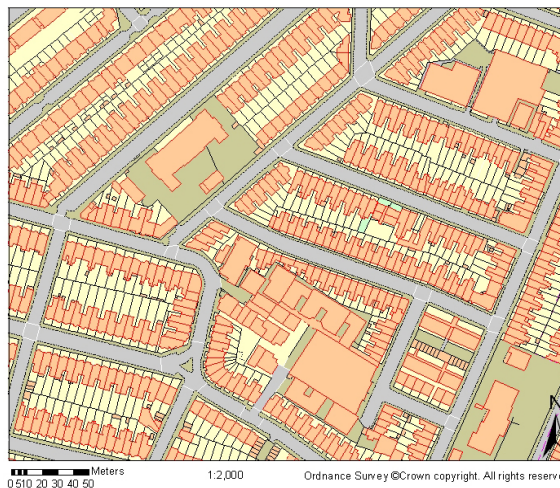
(e) Residential area, London
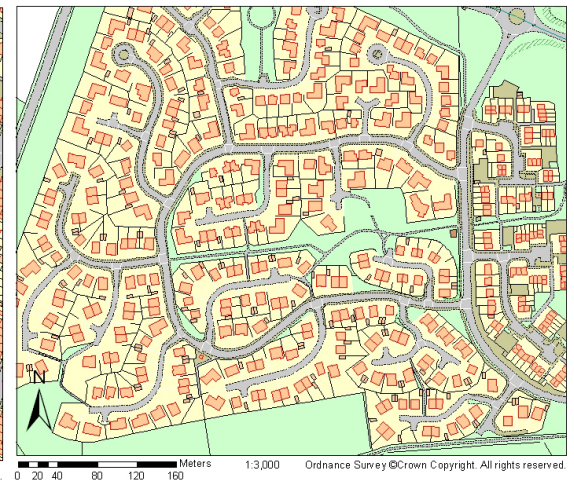
(f) Residential area, Glasgow

Figure 2 Topographic scenes of functional information
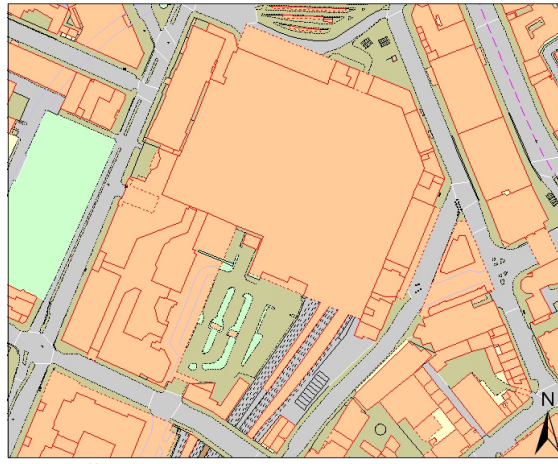
(g) Victoria train station, London
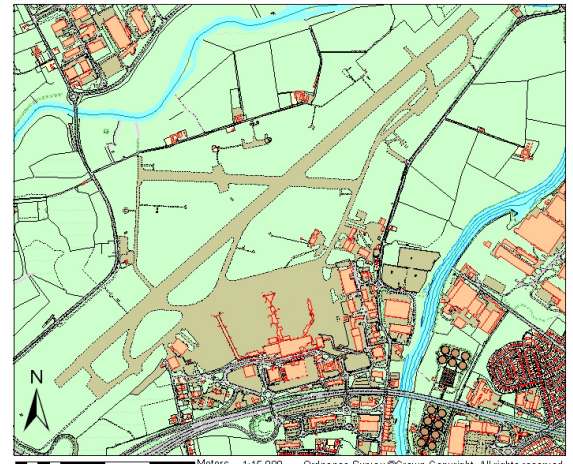
(h) Glasgow International Airport
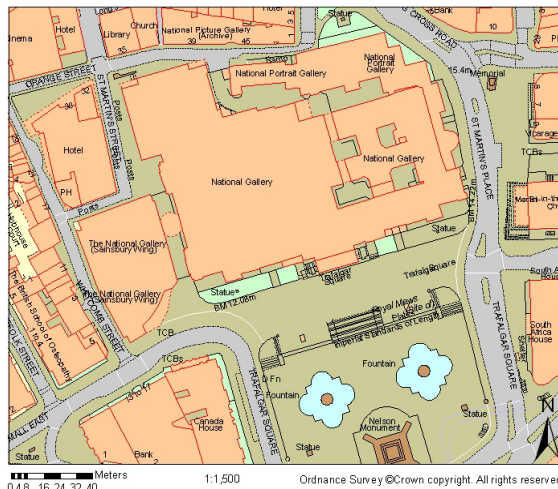
(i) Public spaces, Trafalgar Square in London
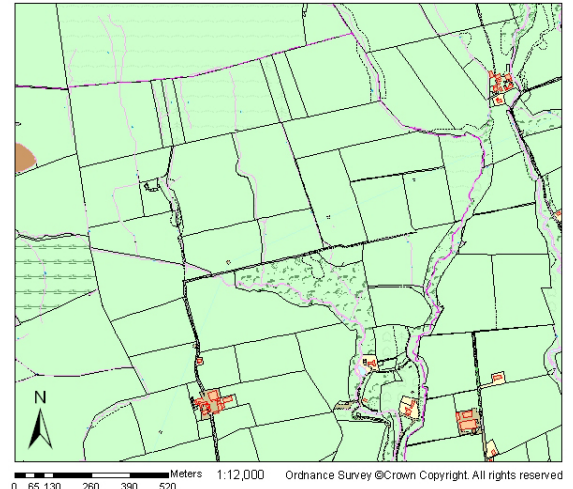
(j) Agriculture, Glasgow

Figure 2 Topographic scenes of functional information

Figures 2 (a)-(j) give only a few key examples for different types of land uses or functions, but the typical organised regularity in the built, man-made environment is clearly evident. Houses spruce up like neatly planted carrots. Roads worm around other features in an organic fashion, connecting them like veins in a body whose blood flow resembles the movement of people in the urban space. Agricultural fields remind one of the patchwork of a large blanket, whose square geometric units can be combined to multiple patchwork blocks, which in turn are joined together to form a larger finished piece, offering almost endless possibilities of different block variations. We could continue to make comparisons of the compositional character of the spatial environment. However, the point is that functional information, although not explicitly, is noticeably represented in spatial data. We therefore need to treat data representation as a compositional problem that consists of a hierarchy of minimal meaningful units,

such as primitive entities (building, roads, etc.), that combine to form higher-level meaningful composites of sets of elements (such as residential blocks). Of course, a detailed data analysis is necessary to identify concrete characteristics of the land use composites, but rich knowledge is present in the aggregation of meaningful object configurations, special relations, perception (Gestalt principles) and context. Because humans can identify such information with ease, we need to achieve a one-to-one mapping between the rich, semantic knowledge and the represented data. This can be achieved by decomposing the rich knowledge to its finest level of detail in terms of its 'syntax', that is, information about the structure of spatial objects (e.g. roads, buildings, land, and water) and how these comprise larger units that convey functional meaning through their spatial relations. Similar to the configuration design problem earlier, we can therefore create a hierarchical configuration where the top-level function is decomposed into its smaller substructures, each of which represents variants of the original goal.

From syntax to meaning

Configuration is a powerful means to say simple things about space and form. Configuration seems in fact to be what the human mind is good at intuitively, but bad at analytically. According to Hillier (1996), we easily recognise configuration without conscious thought, and just as easily use configurations in everyday life without thinking of them, but we do not know what it is we recognise and we are not conscious of what it is we use and how we use it. Apparently, we have no language for describing configurations, that is, we have no means of saying what it is we know. Hillier calls this non-discursivity, and labels it as the central problem of architectural theory.

Interestingly, Hillier uses language as an example, and differentiates between social knowledge (that of spatio-temporal phenomena) and analytic knowledge (that of configurational structures that link words into meaningful complexes). He argues that language only works because we are able to use the configurational aspects of language in a way that makes their operation automatic and unconscious. The words we think of are at the level of conscious thought, while the structures we think with, which have the nature of configurational rules in that they tell us how things belong together, are hidden.

The structuralistic approach, with Ferdinand de Saussure as its originator in the 20th century, focuses on these regularities and patterns, which are a manifestation of the underlying system of language, or semiology (Chierchia, 1999). For Hillier *"structuralism is an enquiry into the unconscious configurational bases of social knowledge, that is, it is an enquiry into the non-discursive dimensions of social and cultural behaviour"* (1996, p.42). He strives to generate and describe the morphology of patterns through the adoption of natural languages and mathematical concepts, thence creating a theory of morphic languages whose syntax captures the elementary objects, relations and operations as realisations of the syntactic structures in the real world (Hillier, 1976).

Language, indeed, offers a powerful ground for comparison. It is a dynamic phenomenon, which varies along the social and gender dimension, from speaker to speaker in idiosyncratic ways, and of course, varies due to a speaker's ability to produce and understand an indefinite number of sentences, while having finite cognitive resources. Yet, naturally, people are endowed with abilities to acquire language and to extract regularities from the environment. Natural languages across the world are built on the same components: a vocabulary (lexicon), a list of words (terms) that are used, a syntax and grammar that describe how valid sentences can be formed from these words, and semantics indicating what the sentences mean (Frank and Mark, 1991; Pinker, 1991). How meaning can be interpreted from a symbolic structure of some kind lies right at the centre of cognitive studies. According to Chierchia (1999), it is our knowledge of meaning that enables us to interpret an indefinite number of sentences, including ones we have never encountered before. The interpretation procedure is therefore compositional: From understanding the meaning of words (or morphemes), to the composing meaning of composed phrases, we cycle through syntactic configurations and arrive at the meaning of the sentence. Our knowledge of sentence meaning then enables us to place sentences within a wider network of semantic relationships with other sentences.

Similar to this, we can interpret our spatial environment by placing the meaning of individual objects into groups of objects, and the meaning of groups into even higher composites of groups, and so on. For example, an individual house can become part of a row of terraces. A row of terraces can become part of a block of terraces defined by

surrounding roads. Blocks can become part of a neighbourhood of terraces, which in turn become part of a complete residential area. Whereas context helps people to evaluate more accurately the meaning of a given area, or group of objects, so does the knowledge of what an object is used for. Especially knowing the function of an individual object like housing will allow us to infer a group of such objects constituting the area's function as residential area (Miller, 1978). Both context and function increase our conversational bandwidth and its richness (Dey and Abowd, 1999). It not only helps us to recognise instances of a category, but to interpret its accurate meaning. As a result, both the ability to express ourselves with language and to convey spatial information (Tversky and Lee, 1998), as well as the underlying ability to interpret meaning from words, sentences and objects, is a fundamental part of this problem. Natural language processing (e.g. Katz and Fodor, 1963; Lewis, 1970; Blutner, 2002) therefore translates to semantic data processing (figure 3).

```
┌─────────────┐
│  Semantics  │
└─────────────┘
   ↕       ↕
┌─────────────┐
│  Concepts   │
└─────────────┘
   ↕       ↕
┌─────────────┐
│    Data     │
└─────────────┘
```
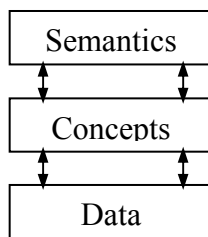
Figure 3 Semantic data processing

The meaning of functional information is a function of the meanings of its parts and of how they are spatially combined. Knowledge of the syntax provides the properties of expression, in this case spatial objects, which are grouped together into hierarchies and at different levels of abstraction combine to parts that in turn create wholes (Chaudhry and Mackaness, 2007). Syntactic knowledge states structural information and relationships about the order of objects in a group signalling a particular meaning. Morphological knowledge relates to the understanding of multiple forms of objects and their spatial order. Semantic knowledge defines meaning about context and how concepts relate to the properties of expression and relations in the world. Consequently, meaning is implied contextually and configurationally, and thus can be exposed by reasoning about the semantics of a dataset although it is not explicitly expressed in the data itself. All that is required is to link the semantics of concepts to the underlying data (figure 3), hence enabling reasoning over that data and thereby making it more meaningful.

2.3 A single, integrated, multi-resolution master database

"DNF will foster an environment where users should not need to capture information that already exists. In future information can be re-used and added together to form new datasets building on existing proven components."

–DNF White Paper (2004, p.14)

Since the advent of relational database systems, spatial data have been managed in database systems. The main application that drives research in spatial database systems is the technology for GIS (Güting 1994). The concern of data storage has now shifted largely to research on human-computer interaction, data sharing and general usability. Concerning issues include spatial cognition, geographic visualisation, multi-scale modelling, and spatial ontology and reasoning. For example, GIS environments and their common users have been studied at the workplace, in schools, and at home to demonstrate the scope of usability issues and the potential of developing techniques and methodologies within this domain (e.g. Haklay and Zafiri, 2008; Davies and Medyckyj-Scott, 1996).

It is important that we apply our experiences from these problems to the development of new systems. Tools and methods should take into account the special characteristics of geographical information and its manipulation, assisting in the design of user interaction. This is coupled with issues in searching and retrieving information from vast amounts of data that are often highly heterogeneous in terms of record types, thematic content, level and type of documentation, and computing environments (Fabrikant and Buttenfield, 1997; Wiegand *et al.*, 2002; Haklay, 2006). However, if the content is not there, or is stored in forms that cannot be converted to those necessary for a given type of model, then any number of research studies and prototype models will fail to result in real-world applications (Davies, 2006).

Consequently, any work that attempts to bridge cognitive science and GIScience needs to consider the relevance of the topic to real world applications, and specifically to the cognitive details of the tasks that people perform. With these issues in mind, how can we improve data accessibility and reuse, why and how is land use information important to users, and how does this type of information make data more definitive and

intelligible? The thesis takes the perspective from Britain's national mapping agency Ordnance Survey to study prevailing issues around data interoperability, user focus and system flexibility, and to gain an understanding of the importance of functional information in the light of these questions.

The need for land use information

Ordnance Survey MasterMap is Britain's state-of-the-art digital map database with over 400 million real world features mapped according to a consistent, national framework being updated continually. OS MasterMap has developed from the Digital National Framework (DNF) project to provide the basis for delineating and maintaining land use parcels. It serves as a reference source to other data through its features' unique identifiers that allow association to related services relevant to the same object. However, the development of a complete land use dataset is more challenging given the lack of existing data sources capable of providing comprehensive information on land use (Harrison, 2002). Most existing information on land use are statistical, hence providing only a general picture of the land use distribution prescribed by administrative districts or wards. For example, on behalf of the Office for National Statistics' Neighbourhood Statistics service, the Office of the Deputy Prime Minister (ODPM) produced generalised land use statistics that cover nine simplified land use themes including domestic buildings, gardens, non-domestic, road, rail, path, greenspace, water and other. These statistic are calculated for each local authority district and each Census ward as defined for 2003, and are provided for all of England as at 2001 (ODPM, 2005). You can calculate land use proportions for areas of interest as percentages, as shown in figure 4.
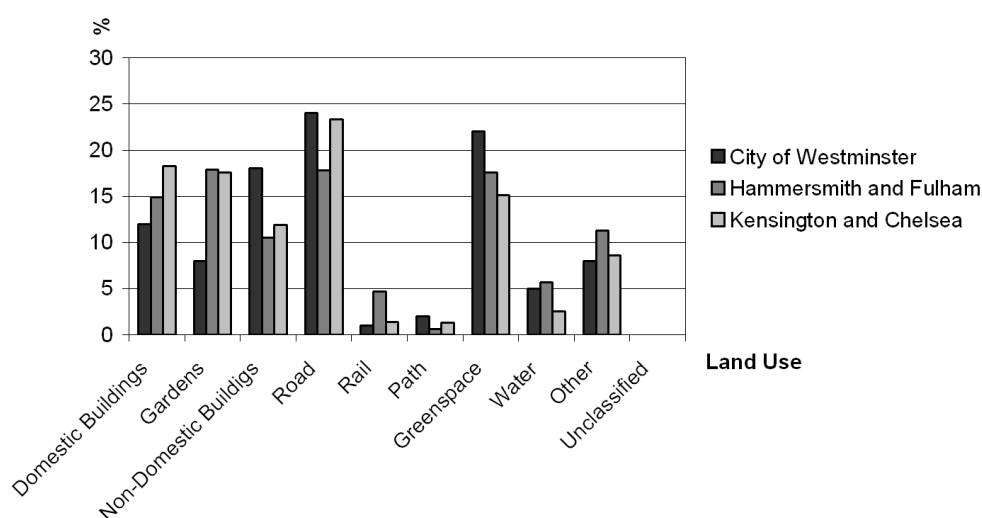
Figure 4 Generalised land use statistics for three London boroughs

The bar chart quickly reveals the highest and lowest portions of land uses across three London boroughs, which are road and greenspace, and rail and path, respectively. You can also use these land use statistics to determine the extent, distribution, and spatial variation of each of the land use categories among different geographical areas (figure 5). By referencing the land use statistics at ward level to the relevant local authority boundaries, which are provided in OS MasterMap, we create a thematic representation of land use per category according to its percentage. This reveals distinct patterns of land use distribution based on defined intervals appropriate to the percentage range of each category.
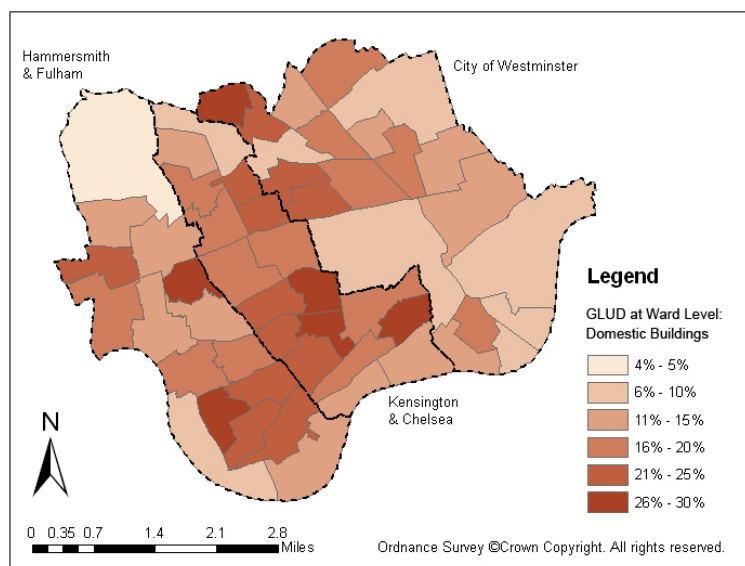
Figure 5 Thematic map of domestic buildings in three London boroughs

These kinds of data, however, are too general for the level of detail required at large-scale representations. In fact, over the past thirty years the development of a standard land use classification and collection of detailed and up-to-date information about the extent and distribution of land use at a national level has failed. Therefore, as described by Tompkinson *et al.* (2004), a series of studies commissioned by the Department of the Environment led to the conclusion that land use should be collected and maintained in collaboration with Ordnance Survey's large-scale digital mapping. In response, ODPM launched the National Land Use Database (NLUD) project in 1998 to develop a comprehensive, complete and consistent source of land use information at the national level based on a standard land use classification (ODPM, 2006). Ever since, many efforts tried to integrate and apply various data sources from the Public Domain to establish such a national land use dataset (e.g. Wyatt, 2002 and 2004; Harrison, 2000 and 2002; Harrison and Garland, 2001; Tompkinson *et al.,* 2004). However, existing

data sources are unable to provide comprehensive information on land use. In addition, automatically generated land use classes suffer from insufficient completeness and confidence values. This meant Ordnance Survey had to reassess the business and technical feasibility of its approach for incorporating land use and land cover data into future releases of OS MasterMap (Harrison, 2002).

Traditionally, map providers have been designing their range of map products to respond directly to the needs of different groups of customers. These products are derived from data that are being collected and maintained involving a large amount of manual work. This often leads to data providers imposing their view of the world, which is traditionally topographic, onto customers who then must modify their understanding of the world to fit that particular view (Byrom, 2003). The contemplation of user requirements somewhat allows to remedy this predicament by moving away from the 'one size fits all' approach towards a 'fitness for use' approach in data provision. Ordnance Survey's key strategies, for example, are to get closer to the customer as part of the renewed customer focus, and to be an 'influencer' and a 'thought leader' of the nation's GI-related activities. "*A key goal is to develop a better understanding of the aims, objectives and applications of our users and customers to further refine the data and information we supply and thereby make it easy to adapt, use and exploit, not only today but in 2008 and beyond.*" (Ordnance Survey Geographic Information Strategy 2006-08, p.3).

Since the late 1960s, Ordnance Survey has positively started to identify and to meet customer needs (Marles, 1983). Instead of falling for the temptation of supplying what has been traditionally provided all along assuming that is what user wants, Ordnance Survey seeks confirmation about current and future needs by consulting users themselves or commissioning professional market research. By visiting end-users in their actual workplaces, and assessing the geographic relevance of their everyday work, it becomes apparent that users do have more to say about the generic and future aspects of their work, as well as trends and events that are likely to change their jobs in future, than is often anticipated. Davies *et al.* (2005) point out that often a product would have been designed fundamentally differently from the start, based on a different set of concepts and structures, if the users' needs and how the product would be used had been precisely understood. In fact, the events in a system's real context of use can vary

significantly from that hypothesized during the design, development and implementation of that system (Davies and Medyckyj-Scott, 1996). However, it requires a lot of effort to discover the true needs of users, as opposed to 'wants' which are endless. *"We want everything really! ... Everyone wants the most detailed information they can lay their hands on."* (Ordnance Survey, 2005-2007). Furthermore, it is evident that users do not fully understand the potential of OS concepts and the data they are using, which leads to an absence of analytical analysis of OS data. Even Davies *et al.* (2005) argue that users often do not know what they want or cannot express it clearly. As a result, the reality looks like this: *"It's not a case of what do I need to do, it's a case of, well, I have got this dataset, what actual use can I make of it."* (Ordnance Survey, 2002, p.12). For this reason, Ordnance Survey conducts internal research to determine the part geography plays in users' lives, that is, what, where, when and how geographical information (GI) is important to them. This enables better decision making for future OS processes, data, products and services, and helps to explore future data needs for GI use in terms of data content and quality (Davies *et al.*, 2005).

Ordnance Survey carried out a project to understand the requirements and information needs of potential users in relation to land use and cover, and to make an informed decision about the development of a land use theme in OS MasterMap (Ordnance Survey, 2002). The study confirmed the general lack of consistent land use data, and that most respondents do not purchase specific land use datasets because few are aware of any that are available. Overall, the research established that there is a wide acceptance and support of the distinction between land use and land cover. Land use is generally regarded as the most valuable because of the number of applications it can be applied to, e.g. government and policy initiatives (crime mapping, regeneration in areas of social deprivation), land planning (urban regeneration), land risk (flooding, contaminated land, waste disposal), lifestyle (commercial development, consumer behaviour), and socio-economic modelling (population migration, health, social planning). However, users are unlikely to purchase land use without the land cover layer because it is also believed to have vital applications. *"People want much more integrated types of information ... to be able to click on an area and know land use"* (Ordnance Survey, 2002, p.16).

Users want intelligence behind the data. They want to distinguish the extent of land coverage through visual examination. This immediacy in visual distinction means that the data are more accessible to those without extensive mapping knowledge. However, the interest in representation is divided between information of larger aggregated polygons and information at an atomic level (individual topographic polygons). Although detail is required in some applications – for example, if you only have the primary use for the ground floor of a building, you are going to miss out on a lot of residential information – data at a granular level potentially leads on to there being far too much information. Generalised data and aggregated information are often easier to interpret and give you a better picture of the neighbourhood. Nevertheless, *the more subgroups you have the stronger your data (but you have to be careful how the data is presented because you can overload)"* (Ordnance Survey, 2002, p.20). The Cities Revealed land use classification, for instance, is too general for users' needs: *"What's the institutional building? A hospital? You need to separate that out because it has implications..."* (p.21). It is clear that users want it all: Current and accurate data with detailed information at the atomic level, but also with aggregated information at more generalised levels of representation. This, however, is far-fetched from the reality. Too much data will mean larger datasets and therefore data management may become a problem. Even if OS MasterMap is used, there is not always the necessary infrastructure to implement it. Often tools such as Google Earth are used instead to get an initial feel for a place, but even this is not always used due to internal bandwidth issues.

The importance of recognising users' longer-term needs, and those of their organisation, could not be clearer. Ordnance Survey continuously invests in researching these needs. Its recent Future Users Research, which took place between 2005 and 2007, reveals a lot more about data issues and the types of functional information that would add value to its data. OS assessed the work of its key users (figure 6) to understand how geographic information interacts with other information and knowledge from other sources (Ordnance Survey, personal communication, March 2006). The study focuses on the features and aspects of the British landscape that are important to the customers' job. It offers the customer an opportunity to tell OS how their work and information needs may change over the next few years. The research consists of fifty-six anonymous user-task interview records, which are categorised according to supertask profiles such as flood risk assessment, catastrophe modelling, urban design, transport and network

management, to name a few. Even years later following the land use study, interviewees express similar feelings (Ordnance Survey, 2005-2007):

> *"It would be good if all geographical information required was available to the user from the same source."*

> *"We want to be able to share information with other people more easily."*

> *"In an ideal world the team would have access to a fully integrated, up-to-date and reliable information system for UK data. A free, nationwide information portal... providing up to date, quality checked, reliable information (on e.g. habitat information, geology etc) for use in decision-making would be desirable. This would save many users a lot of repeated work."*

> *"The Ordnance Survey provides some documentation but doesn't, for example, provide easily available definitions for things... Datasets from some other organisations, containing geographic features, may have even less definition for what is meant by terms such as 'lake' etc in the dataset."*

> *"Automated generalisation tools, to consistently generalise datasets from their highest resolution base data could improve the overall quality of data worked with, especially if generalisation could be done on the fly."*

Overview of Customer Segments

```
1        PSU*¹  ┬── Central Government ── PGA*² ──┬── DfT
                │                                 ├── Defra
                │                                 └── ODPM
                │
                ├── Local Government ── MSA*³ ──┬── Local Authorities
                │                               ├── County Councils
                │                               └── National Parks
                │
                └── Utilities ──────── NJUG*⁴ ──┬── Water
                                                └── Electricity*⁵

2        Partners  ┬── Partners
                   ├── System Integrators
                   └── Developers

3  Commercial Markets ┬── Insurance ──┬── Commercial Property Insurance
                      │               ├── Retail/Personal Insurance
                      │               └── Fraud detection
                      │
                      ├── Transport ──┬── Routing & Navigation
                      │               └── Logistics & Fleet Management
                      │
                      ├── Wireless ──┬── Business (Tracking)
                      │              └── Consumer (A-to-B)
                      │
                      ├── Land & Property
                      └── Housing Associations
```

*¹ Public Sector & Utilities *² Pan Government Agreement *³ Mapping Service Agreement *⁴ National Joint Utilities Group
*⁵ Private but linked to government

Figure 6 Overview of Ordnance Survey's customers

The predominant need for data interoperability, to have all the data available in one source with sufficient semantics, and on the fly aggregation is evident. However, so are the information, such as real world things, or concepts, and key attributes, required for the tasks that users have to carry out. In particular, functional information is a versatile component that fits into every task description. *"There is much need for land use data. Anything that can be produced on land use (per field polygon rather than as a grid) would be very useful"*. This is a typical reaction by respondents. Interesting is an analysis of the interview records in regards to what kinds of functional information are valuable. There are too many to mention them all here, but they range from detailed and specific land uses, such as hospital, airport, pharmacy, pub, school, train station, or golf course, to generic types, for example industry, retail, town centre, residential, agriculture and conservation areas. *"For emergency planning generally, great levels of details for buildings etc. are not needed (do not want too much clutter)"*, whereas for other tasks *"many of the other categories are too vague to be useful. [...] In terms of land use, much effort is currently put into identifying green spaces with public access, e.g. parks in towns which aren't necessarily identified as such in current OS MasterMap."* Building use (residential, commercial, etc.) and building type (e.g. detached, terraced, bungalows) are of particular interest, and buildings are considered part of land use. In addition to land use type, the extent of urban areas and settlement

patterns is often needed. This includes building complexes such as airports, stations, etc. where the spatial relationship of buildings to other buildings and features around the property (e.g. roads, schools) are important. With the more general categories, for instance industrial area, it is also useful to identify the type of industry.

The functional geography is complex, as can be seen from the multiple land uses required at varying levels of detail. Figure 7 illustrates an example of such different classification levels. To accommodate all these needs and cater for a greater range of services to customers, Ordnance Survey is implementing a single, integrated and unified large-scale master database from which all future products will be derived. This has implications for the storage of data and its attribution. Functional information will be integrated in the form of the attribute 'base function', which in essence represents the action, purpose or role for which a thing is specially fitted or used. In addition, functional sites will be created, where more than one feature are used together to support or perform a given function. Currently, the attribution is far from complete. 'Base function' is extracted from existing cartographic text until functional information is collected directly when features are being surveyed. The mentioned ongoing projects both internally and externally as well as the user requirements not only emphasize the relevancy and importance of this thesis, but impose implications for a solution, as discussed in chapter 5. With functional information being crucial for so many applications, deriving new information at a low cost solution from existing data remains highly beneficial.

LEVELS ⟶ 1 2 3

LAND USE

Residential
- Bungalow
- Terraced
- Detached
- Flats
- Villa houses

Commercial
- Shops
  - Hairdresser
  - Newsagent
  - Supermarket/Grocery
- Banks
- Clubs/Pubs
- Shopping centre

Industrial & Business
- Development site
- Cental Business District
- Corporate — Office

Utilities
- Street furniture — Post box
- Flood defence structure
- Drains & gullys
- Control & access points
- Telecommunications

Transport
- Vehicle storage — Emergency depot
- Waterways — Canal
- Terminals & Interchanges
  - Harbour/port
  - Bus stop
  - Airport
  - Rail station
- Transport Tracks & Ways
  - Public footpath
  - Bridleway
  - Cycle way
  - Grinnel
  - Pedestrial crossing
  - Pedestrian zone
  - Promenade
  - Footways
- Car parks

Recreation & Leisure
- Conervation site
- Caravan & Camping site
- Resorts
- Amenities
  - Beach hut
  - Beach
  - Baths
  - Pier
  - Social venues
  - Football grounds
  - Libraries
  - Town hall
- Public buildings/spaces
- Parkland
- Landmark building
- Historic site

Community services
- Community centre
- Fire station
- Police station
- Ambulance station
- Post office
- Education
  - Primary school
  - School
  - Campus
- Health & Medical services
  - Hospital
  - GP surgeries
  - Sanitarium

Military
- Armed forces bases
- Battle field

Agriculture
- Farm buildings
- Cattle grid
- Farmland

Unused land — Derelict land

Figure 7 Hierarchical representation of land uses

On-demand mapping

Geospatial data are only as useful as they are accessible, which means the data must contain all the relevant information for a given task, in a format that can be easily shared, analysed and commented on, and at variable scale representations. For national mapping agencies for example, multiple scale map production is a crucial element in the business to derive products with as much automation and flexibility as possible. A limited number of fixed scale maps usually accommodate the different levels of details. These however entail considerable leaps in detail from one scale to another, because at present there are no automatic facilities to modify the level of generalisation of available geo-data in commercial GIS. The popular aim of national mapping agencies is to build a large-scale digital database from which medium- or small-scale cartographic models can be derived (Lee, 2004). The principle is that a multi-resolution spatial database is used to store low-level geometry that is attributed with scale-specific data that enable lines or polygons of a required level of detail to be reconstructed from their component vertices (Jones and Ware, 2005). A system that tailors products according to the end user requires such flexibility, along with the ability to combine different sources of data, as well as an understanding of what the user wants to have in the product and how it is represented. However, current map production systems involve a large amount of manual work, which limits the possibility of producing more custom orientated products.

The problem is that general-purpose topographic databases are poor in semantics. This especially concerns the representation of higher order semantic concepts that extend beyond the meaning of individual, discrete objects. Generalisation rules refer to such higher-level concepts in the form of the spatial organisation, or context, of these objects. The semantic characteristics of map objects are necessary to obtain priority orderings among map objects and to form meaningful groups before informed decisions can be made about generalisation (Neun *et al.*, 2004). What is required, therefore, are methods that make explicit the spatial relationships and semantic concepts implicitly contained in spatial databases (Lüscher *et al.*, 2008). Indeed, the challenge in developing generalisation solutions roots from the complexity of generalisation tasks itself, where no features should be generalised in isolation (Lee, 2004). By enriching the source data with higher-level concepts, these concepts are then available to provide contextual

information for a generalisation process or they can be used as a map component in the final product.

This process of map generalisation is termed semantic generalisation. It is the choice of the appropriate categories of information, or concepts that should be represented. Concepts can describe, for example, patterns and their roles at varying levels of representation. Mackaness and Edwardes (2002) argue that the combination of various patterns enables the creation of higher order phenomena – such as land use. By decomposing and defining patterns in terms of individual map objects, we can associate a set of generalisation methods used to manipulate those objects with the representation of an higher order phenomenon. An automation of the generalisation process then requires a model of these patterns and their transitions in a more meaningful and explicit way (Edwardes *et al.*, 2005). Here, knowledge representations can be particularly useful in providing the necessary structures to model the categorisation of concepts hierarchically. At the top of the hierarchy are the most important concepts, whereas their more specialised subdivisions reside at progressively lower levels. The meronymy of concepts describes the partonomic relationships between objects. Hence, a semantic model essentially defines objects, relationships among objects, and properties of objects. These types of relationships are relevant to guiding semantic map generalisation, whereby finer or coarser distinctions are made between concepts according to the level of abstraction that is appropriate (Jones and Ware, 2005). Consequently, such a model provides a number of mechanisms for viewing and accessing the database schema at different levels of abstraction (e.g. Ram, 1995).

The idea behind a multi-resolution database system is to take some input data, some target specifications (i.e., map specifications) and automatically trigger the sequence of generalisation tools that will transform the input data into the specified output data (Regnauld, 2008). This requires a framework for not only defining conceptually geographic entities and their target specifications of the final map product, but also suitable tools for retrieving and displaying objects in a given context. Geometric generalisation refers to the simplification of the shape and structure of the graphical symbols that represent individual features (Jones and Ware, 2005). Semantic generalisation therefore dictates to some extent the type of geometry necessary to construct map symbols. However, often the data enrichment algorithms themselves are

buried inside a generalisation process resulting in a typical black box approach. The enriched components are consequently dependant on specific implementation characteristics, which are inaccessible to the user, as well as on a particular data model and even schema (Regnauld, 2008). Because generalisation rules are loosely defined, often providing mere guidelines, it would be better to formalise the definition of the higher-level concepts and their derivation rules to allow the user to tweak them.

We need to divorce ourselves from the black box processes to derive different components of the map. Instead, we should explicitly model higher-level concepts, such as functional information, and their associated abstractions. We need to understand individual objects, such as a building, in their context of use. We may want to know which areas of the city are residential, commercial, or industrial, and what spatial extent they have to inform generalisation processes or produce content specific maps. By making these types of high-level information available as components, they can be exposed to whoever creates new products. Ultimately, a user could have access to a structured library of such components to select one or more components he or she is interested in, and then be presented with a list of abstractions available for this or another concept.

This semantic-based approach has the potential benefit of ensuring that concepts are not affected by changes in the structure of the source data, and that different representations of the same concept will be consistent, because the enrichment process is explicit. This will enable different application developers to share the tools processing the data (Regnauld, 2008; Neun *et al.*, 2006; Edwardes *et al.*, 2005). According to Ordnance Survey, it is important to integrate what has already been done, thus reusing existing methods and developing the missing components to deal with new requirements. The issues of generalisation therefore must be tackled in a collaborative manner, for example as part of the ICA commission on generalisation and multiple representations, to avoid temporary in-house formalisms that package and describe software components, which will have to be abandoned in the future to take advantage of a richer source of reusable components. An explicit modelling promises a solution to these problems, including data interoperability, user focus, and system flexibility.

Conclusions

In this chapter, we have learned about the importance of enriching spatial data with higher-level information, or knowledge, to accommodate users' requirements for integrated data at multiple levels of representation. In particular, there is a substantial demand for land use information by different applications and users, as indicated by Ordnance Survey's customer research for incorporating land use into future releases of OS MasterMap. Land use information is regarded as a higher-level concept that is implicitly present within the spatial configuration of features stored in a spatial database or a remotely sensed image. If we revisit the first research question of this thesis, that is, what can spatial context and its configuration tell us about the functioning of its features, then we can conclude that there is indeed an eminent relation between the spatial form and its function. This relationship has been discussed both in its theoretical and practical implications. With the result that many models proposed in urban studies, despite different analytical and scientific views, acknowledge that there is a common understanding of the complex nature of the urban fabric and its functioning. Form reflects function and vice versa.

Hillier's configurational theory seems to offer the key to understanding urban phenomena both material (i.e., form) and immaterial (i.e., function). Many different implementation methods adopted his theory, from graph-theoretic approaches, or rule-based aggregation, to clustering and other classification procedures. Each one tries to create a mapping between land cover parcels and higher-order meaning of the scene such as land use. Despite the common demand for inferring higher-order meaning from spatial data, existing approaches suffer from considerable limitations that are often reflected in the classification accuracy, complexity, processing time, and lack of human capacity of interpretation. According to Minsky (1975), large amounts of knowledge are required to make machines intelligent and to provide intelligent information processing – indeed, *'you cannot tell you are on an island by looking at the pebbles on a beach'*. Interpretation, or inference of higher-order meaning, is a knowledge intensive task, and it has therefore been widely acknowledged that research on information extraction must consider primarily the semantics of the data.

Semantics is often linked to different conceptual levels of abstraction, as higher-level information is largely described by meso objects – a description of a combination of individual, micro objects. If we consider built environments as organised systems whose primary nature is configuration, then we must treat inference as a configuration problem. Instead of building up a configuration from individual components, we need to decompose the high-level entity into its constituting elements. A given land use type therefore specifies all its necessary components such that a residential area, for example, consists of residential houses in terraced, semi-detached or detached form. We need to identify the requirements that specify what types of land cover parcels constitute a land use and constrain it through their general morphological properties. Configuration problem solving then consists of finding the optimal layout in the data and aggregating individual objects into meso structures. Of course, land uses cannot always be neatly categorised. In the real world, one will find that more than one land use can exist for the same parcel of land. In figure 7, for example, land uses at the very fine level of detail (level 3) are too specific to be inferred from land cover data alone. These kinds of information are often available in points of interest data or as cartographic text labels. However, the general top-level categories (level 2 and 1), such as residential or industrial, are more likely to be implicitly stored within spatial data because they form the conceptual aggregation of individual, discrete objects. Specific land uses then can be aggregated into their primary, general land use of a given area.

Consequently, we need to create a one-to-one mapping between rich, semantic knowledge on the one hand and the constituting syntax of land cover objects on the other hand. The interpretation of the spatial environment is then achieved by placing the meaning of individual objects into progressively, higher-level groups of objects, similar to the way we process natural languages. By exposing these higher-level structures, data will reflect more the way people perceive the world, not just the geometry of physical topographic features (Montello, 2002). Knowledge representation formalisms offer useful structures for modelling these kinds of high-level concepts. This thesis will explore a knowledge-based approach to innovate the area of inferring land use information directly from topographic maps. First, however, we need to learn about the types of processes and knowledge involved in such a task.

Chapter 3

Interpreting Higher Order Meaning from Topographic Maps

"One reason for the deficiency in the representation of geographic phenomena in a way that is appropriate for a wide range of application contexts is that the conceptual models currently employed for such digital geographic data representation do not incorporate any explicit consideration of how humans cognitively store and use geographic knowledge."

–Mennis *et al.* (2000, p. 501)

We use and live in our environment taking it for granted, acting upon it almost unconsciously, exploiting it, referring to it, solving our daily spatial problems of how to get from one place to another and absorbing a daily wealth of information. Yet, we do not realise the wonder-like capabilities that are tied with these processes that seem to happen so effortless when performed. I am referring to our so-called black box of a mind that allows us to reason, acquire and structure our knowledge; let alone to which nouns like personality, thought, memory, intelligence and emotion are subscribed. Indeed, a variety of social sciences especially kinds like psychology and anthropology are concerned with discovering the mind and its fundamental cognitive processes. Here, however, a pure geographical perspective is taken to investigate more closely our spatial knowledge of the functional environment and how we interpret its spatial characteristics as depicted in maps.

The most universal and well-known representation of geographic phenomena is the map. The map is special because it is both a graphic image as well as a geometric structure in graphic form (Peuquet, 1988). The variation of lightness and darkness, pattern, and possibly variation in colour characterise the map as an image. It may or may not convey meaning, as in the case of an abstract painting. The map's geometric structure, on the other hand, provides an unambiguous representation in an appropriate coordinate system. Peuquet (1988) argues that since maps are human-derived representations of geographic space, this image versus structure duality also

holds for how humans perceive geographic space, corresponding to the world as seen (image) and the world as understood (structure). In models of human perception of the spatial world, it is generally agreed upon that there is a distinction between what is seen and what is understood. In fact, what is seen is the result of a synthesis of different types of input: visual, auditory, olfactory, and kinaesthetic (Downs and Stea, 1977). Some may argue that people perceive their environments in similar ways because of these physiological similarities (Mark and Frank, 1996). Others believe that the world is perceived in an individual way. Neisser (1976), for instance, claims that people only see what they know to look for, what they expect to see, and what they want to look for. In contrast, what is understood from the perceived world must be interpreted based on prior knowledge or experience, which in turn can also become knowledge and influence how subsequent inputs are interpreted (Peuquet, 1988). *"Perhaps much of the confusion that lies at the heart of geography today results from an awareness that there are simply many geographies and many possible worlds"* (Golledge, 1982, p.21). Indeed, no single representation of the world incorporates every possible viewpoint. The myriad of geographic models mirrors this dilemma to the degree whether a small-scale space relates to body sizes and smaller (Siegel, 1981), or to that of a single room (Gärling and Golledge, 1987), or to that of a tabletop view of a large space (Mandler, 1983). Hence, the discrepancy of space representation relates to both the psychological connection to the world as well as to geographical and mathematic representations of space (Mark and Frank, 1996).

This dichotomy reveals the importance of bridging the gap between widely deployed models of space and what research in cognitive sciences identified as being important for human interaction with and conceptualisation of space (Mark *et al.*, 1999). For example, Mennis *et al.* (2000) point at existing conceptual models employed for digital geographic data representation, which do not incorporate any explicit consideration of how humans cognitively store and use geographic knowledge. It is important that spatial data models represent information in a way that is more natural to humans. This will not only result in improved spatial information processing, but also accommodate a wider range of application contexts. To get a step closer to filling this gap, we need to ask specific questions about people's understanding and reasoning over space, the environmental characteristics that influence people's knowledge, and how all these aspects can be modelled in a way appropriate for the

computer to process this information. Therefore, this chapter addresses two important issues. Firstly, we need to learn more about people's perception of space and spatial cognition to define an appropriate solution for better information processing and inference. Secondly, because interpretation is a knowledge-intensive task, we need to find relevant persons to gather their reasoning skill and specialised knowledge – intellectual cloning as Wilson and Keil (1999) call it – to embody that knowledge and reasoning later in a program, or other form accessible to the computer.

Social sciences apply both qualitative and quantitative survey techniques, such as interviewing or questionnaire research, to unravel the unknown from human behaviour, whether it is exploring reasons behind processes or relationships among phenomena (Pope and Mays, 1995). A questionnaire survey thus provides the relevant material by aiming to derive consistent cognitive information from human experience of geographical space (Thomson, 2007). Questions are posed about the types of structures and patterns perceived according to each land use concept, the relevant (cognitive) approaches for identifying land uses within a topographic map, as well as on the common properties for the instances of a concept and the spatial structure among these concepts. By dissecting this information, the spatial environment is cut up and organised into concepts, and knowledge is induced. This chapter therefore analyses the problem of interpreting land use information from a human perspective. Together with relevant theoretic underpinnings and existing experiments as described in the next section, the questionnaire survey will indicate necessary knowledge and reasoning skills required for deriving a solution (section 3.2).

3.1 Spatial perception, cognition and categorisation

Human cognition takes place in a social and cultural context making use of tools such as language and communication, concepts and beliefs. Arguably, the very existence of culture is both an effect and a manifestation of human cognitive abilities, and human societies of today culturally frame every aspect of human life and cognitive activity (Sperber and Hirschfeld, 1999). Many different forms of cultures have evolved over time, perhaps not as an effect of biological variation, but more specifically from cognitive endowment given that different historical and ecological conditions make such variations possible. For example, studies of folk biological knowledge and their

classifications postulate that knowledge is based on a domain-specific approach to living things characterised by specific patterns of categorisation and inference (Atran (1990). Even though the initial approach to distinguishing these domains may be general across cultural knowledge, domain-specific knowledge results from the variation of identification and interpretation of such phenomena. Thus, despite the disposition to classify animals in the same way, local faunas differ and so does people's involvement with this fauna (Sperber and Hirschfeld, 1999). As a result, domain-specific abilities contribute to explaining cultural diversity, as the information processed meets specific input conditions that depend on the environment. These environments are not always natural, and the greatest variation of cognitive disposition is found across our artificial, man-made environments. Hence, culture may be thought of as an ensemble of representations or classifications, schemas, models, etc., whose possession make one become a member of a cultural group. Its pool of traits enables, constrains, and channels the development of cognitive outputs. For instance, different language systems have an impact on segmentation, categorisation, and modes of thought in general, hence limiting our abilities to conceptualise the world (Tversky and Lee, 1998; Berlin and Kay, 1969). This means that not only do human mental processes make use of cultural tools, like language, models, expertises, values, etc., but they are also a reflection of that culture, just as they are a reflection of the environment at that place and time. As a result, everything that surrounds us shapes our knowledge – let it be socially, culturally, experientially or environmentally.

Our interaction with the environment is a correlation between what is perceived and how the perceived is interpreted. Max Wertheimer (1924) established the most predominant principles of perception, and became renowned as founder of Gestalt theory. Gestalt theory overrides the previous perceptual theory of a structuralistic nature by arguing that people perceive organised scenes consisting of surfaces, parts, and whole objects coherently arranged in space rather than as a chaotic, dynamic juxtaposition of millions of different colours registered by retinal receptors. Stimulus factors, such as proximity, similarity of colour and size, common fate, good continuation, common region, closure and element connectedness, cause elements to be perceived as organised in distinct groups. For example, shape is related to concepts such as form and structure, and provides valuable clues about an object's identity, as

well as information that are critical to manipulating objects and determining their functional properties or affordances. Shape perception depends on part on feature extraction processes, and processes that group elements into higher-order units. We can perceive both the shape of the individual elements, as well as the global shape of the grouped elements (Wilson and Keil, 1999). This emphasizes the doctrine of holism based on the assumption that the whole, i.e., the group of elements, is different from the sum of its parts, or individual elements.

People's awareness of space is a result of how we explore geographic space by navigating in it, and how we conceptualise it from multiple views, which are put together mentally like a jigsaw puzzle (Egenhofer and Mark, 1995). Golledge (1992) examined in detail the components that make up spatial knowledge: The location of occurrences, spatial distribution of phenomena according to pattern, shape or density, regions or bounded areas of space, hierarchies, networks, spatial associations, and surfaces or generalisations of discrete phenomena. Many researchers draw an analogy to the cognitive map with which they metaphorically describe how people process and recall spatial information (Holyoak, 1999). We retrieve knowledge according to conglomerations of information drawn from different sources and modalities that are pulled together for a particular purpose or problem-solving task (Mark *et al.*, 1999). For spatial problems, such as navigating through the environment, we rely on our sophisticated mental representations of spatial relations. The cognitive map is one way to describe the mental representation that is derived from the environment and allows us to make sense of that environment.

There are numerous metaphors out there each trying to describe more appropriately the mental processes that lie beneath our spatial knowledge: From the cognitive map (Tolman, 1948), imaginary map (Trowbridge, 1913) spatial images (Lynch, 1960), cognitive atlases and the 'map in the head' metaphor (Kuipers, 1982), to spatial mental models and cognitive collages (Tversky, 1993). The image comes much closer to what a map in the mind resembles metaphorically. A picture is worth a thousand words (Pinker, 1997), and as such ideally serves as a means to achieve cognitive economy. However, not any two mental representations can be similar between two people. Mental representations, or cognitive maps, are greatly influenced by experience, age, and styles of training and thinking (Downs and Stea, 1977). From a

geographical point of view, a mental representation can never exactly portray its captured spatial environment. Previous psychological research has revealed that spatial information is largely distorted and subject to systematic error (Tversky, 1993; Lloyd and Heivly, 1987). Furthermore, humans are lacking the storage capacity to allow perfect identity between representations and the spatial environment (Downs and Stea, 1977). This reflects somewhat Alexander's point (1965) that the limited capacity of the mind to form intuitively accessible structures results in simplified representations that cannot encompass the complexity of the real world in all its facets.

Corresponding to psychologists, the purpose of this mental mechanism is to cut down information, or to generalise it into manageable portions for memory to hold all the events. By recognising, differentiating and understanding it, we categorise the conceived information into apprehensible chunks, assigning categories to one class according to shared characteristics, and across classes based on their distinct characteristics. Eleanor Rosch (1978) revealed that categories arise out of an interaction between stimuli and process. This means category processors, like human beings, require the ability to judge similarity between stimuli, to perceive and process the attributes of a stimulus, and to learn. On the one hand, the function of the category system, or classification, provides maximum information with the least cognitive effort, thus aiming for cognitive economy. On the other hand, the perceived world comes as structured information rather than as arbitrary or unpredictable attributes. If categories map the perceived world structure as closely as possible, then also a state of cognitive economy can be achieved. According to the latter principle, the category system is already existent in the culture at a given time.

Rosch (1978) proposes a category system based on vertical and horizontal dimensions. In general, the vertical dimension refers to the level of inclusiveness of the category, whereas the horizontal dimension refers to the segmentation of categories. This has the implication that in the vertical dimension not all possible levels of categorisation are equally good or useful. According to Rosch, the most basic level will be the most inclusive level at which categories can mirror the structure of attributes perceived. The basic level is the one first learned by children, preferred in naming and most rapidly categorised by adults (Wilson and Keil, 1999). In the

horizontal dimension the distinctiveness and flexibility of categories is increased. Where a category member is placed within the structure depends on its typicality, the degree of category membership, or goodness of example. The more prototypical of a category the member is rated, the more attributes it has in common with other members of the category and the fewer with members of contrasting categories. Hence, separateness and clarity of continuous categories is achieved by conceiving each category in terms of its clear cases rather than boundaries.

A category member equals a concept or object whose belonging is determined by its shared attributes or characteristics. In Wilson and Keil (1999, p.176) the term concept is defined as "*the elements from which propositional thought is constructed, thus providing a means of understanding the world, concepts are used to interpret our current experience by classifying it as being of a particular kind, and hence relating it to prior knowledge*". Various views of the concept exist, whose elements are, according to Smith *et al.* (2005), often found mixed up together in almost all terminology-focused work in informatics nowadays (table 1).

Table 1 Views of concepts

View of concepts	Definition of concept
Psychological view	Mental entities, analogous to ideas or beliefs
Linguistic view	Meanings of general terms
Epistemological view	Units of knowledge, as in knowledge representation
Ontological view	Abstractions of kinds, attributes or properties
Theory of basic levels	Basic level of categorisation corresponding to high-frequency used nouns

From a geographical point of view, concepts and objects must form some relation between the geographical world and our understanding of it. Following Wüster's definition (Smith *et al.*, 2005), an object is defined as anything perceived or conceived: Some objects are considered as material, some as immaterial or abstract, and others as purely imagined. Consequently, it can be said that geographic objects depicting real world spatial entities are materialistic things, such as a river or building, whereas their perceived attributes, which are the ways in which humans habitually use or interact with those objects, are immaterial or abstract 'objects' in that sense. With respect to Rosch's theory of basic levels, there should be a basic level of geographic objects. Similarities of the prototypes and structure of such a basic level is found in

some categories describing geometrical forms (for example, circle, square and equilateral triangles), and judgements of physical distance (Vorwerg and Rickheit, 1998).

Considering these aspects, it is important to find a coherent means by which concepts and their characteristics can span the divide between concepts as creatures of the mind and as properties of objects in the world. To build a conceptualisation worthy of representing our geographical environment according to its interaction with human beings requires a bridge between human knowledge and that of real world entities. Although, the lingering incoherence is reflected by manifold representations of ideas in people's minds, meanings of words, and consensus knowledge of experts in a discipline or types of entities in the world (Smith *et al.*, 2005). We can expect from the above account that the conceptualisation will be more or less an exact reflection of the real world. If the category system is already existent in the culture at a given time, then how differently, if at all, is this information organised in our heads? More importantly, how can we accurately capture this structure and its containing knowledge? According to Mark and Frank (1996), mental models can reveal themselves through spatial reference in natural language, through experiments with human subjects, through observation of spatial behaviour, or through study of the artefacts of such behaviour. The form of mental models is expressed in either of two media, imagery (mental maps) or words (categories) (Downs and Stea, 1977). Yet, Harding and Davies (2004) claim there remains great uncertainty as to the 'best' model for human categorisation even in relatively straightforward domains such as biological kinds. How can a conceptualisation that represents geographical as well as functionally abstract concepts account for fuzziness, for example? Crispness may only exist in idealisation or system of rules, which abstracts away from complicating aspects of reality, as Pinker (1997) claims. Yet, a concept becomes fuzzy again if taken out of its idealised theory. People form concepts that find clusters in the correlational texture and vicissitudes of the real world. As Rosch (1978, p.42) notes, "*it is predetermined that there will be context effects for both the level of abstraction at which an object is considered and for which items are named, learned, listed, or expected in a category*".

Conceptual encoding, or conceptualisation, thus occurs in a wide range of phenomena in perception, attention, language comprehension, and memory. For example, concepts help us to grasp and communicate about spaces. The way location is described by a formal model through coordinates or its scientific geographical vocabulary, differs largely from the way people describe a location on a day-to-day basis (Hockenberry, 2004). These descriptions not only identify places, but give members of our socio-linguistic group information about them, building up a jointly-defined cultural world-view within which we all act on a daily basis (Waters and Evans, 2003; Agarwal, 2004). Conceptualisations, however, vary greatly among different domains. For instance, the conceptualisation of topography varies greatly among topographic cartographers, information scientists, and geomorphologists, between that of pilots, explorers, anthropologists, hikers, and archaeologists (Mark and Smith, 2005). This yields the question what people will look for in a topographic map, depending on their knowledge and expectations, and how principles of perception will guide people's understanding. Those discrepancies highlight the importance of incorporating cognitive aspects into our geographical models. And indeed, the nature of geographical knowledge and its research has been changing; changing from its declarative nature of collecting and representing the mere physical and human facts of existence, to the creation of knowledge generated by emphasizing cognitive demands focusing on processes and asking the 'why' and 'how' questions (Golledge, 2002).

3.2 Relating land use to the landscape character: An investigation

Questionnaires provide an objective means of collecting information about people's knowledge, beliefs, attitudes, and behaviours. They offer a useful instrument for capturing knowledge as part of ontology engineering (Thomson, 2007), which is concerned with formalising a conceptualisation of a specific domain. Prior to any conceptualisation, a recording of the relevant knowledge is necessary. Finding a source of expertise that can be harvested is one challenge, the other is faced by how this information can be most efficiently extracted. The relation between humans and their knowledge about and their interaction with the environment is not an easy one to ground, as can be seen from numerous theories such as empiricism, positivism, rationalism, idealism, or constructivism, that offer different explanations for the

nature of knowledge (Gale and Golledge, 1982). Knowledge is something conscious, something that needs to be recognised. As Downs and Stea (1977, p.xiii) note: "*The ability to understand ourselves is limited by the ever-present feeling that much of how we behave and how we think is obvious, that these are things that everybody 'knows'. We hamstring ourselves with the disparaging remark that something is trivially obvious.*" Often knowledge, or lack of knowledge thereof, becomes only apparent when it is required, such as when we are taken out of our familiar environment and loose sense of orientation.

From a geographical perspective, knowing is constituted by the environmental habitat that provides the necessary context for learning with constant feedback and adjustment (Hutchins, 1995). Indeed, knowledge can be gained by direct experience, but as well by facts and second-hand information. However, what is individually understood as knowledge depends largely on a person's beliefs and truths that one confides in. Beliefs are thought of as psychologically-held understandings, premises or propositions about the world that are thought to be true (Richardson, 1996; Hofer and Pintrich, 1997). Therefore, what is accepted as knowledge may well be infiltrated by certain beliefs and truths that may lead to 'contaminated' knowledge. In respect to specific technical knowledge, this may not be such a concern, as knowledge is well documented and can be gathered from a domain expert.

Most knowledge acquisition techniques focus on interviewing a single domain expert, who might be directly involved in the project itself, or extracting knowledge from loosely structured textual or multimedial data, or databases (Svátek, 2006). However, if the system relates to the geographical domain and its rich yet familiar phenomena, then subjects become malleable to its physical, cultural and social influences that provide people with the information to be perceived, processed and conceived, as we saw earlier. Thus, beliefs vary according to gender, ethnic, cultural difference, and spatial context. Nevertheless, geographical knowledge is concerned with common or natural phenomena. Since we are dealing with human lives and their perception of the environment that poses as the normal setting for people's activities (Downs and Stea, 1977), every person becomes an expert in their own right: "*Anyone who inspects the world around him is in some measure a geographer*" (Lowenthal, 1961, p.242). If knowledge is to be drawn from a wider population, then survey techniques as

employed by social sciences become a means to an end. After all, it is the knowledge derived from a number of experts that will always exceed that of a sole person allowing for the generalisation of the subjective to the objective interpretation of knowledge.

Aims of the survey

If we recall, the overall aim of the thesis is to derive a mechanism that can (semi-) automatically process a topographic database to infer additional, previously implicit information of a functional nature – something, so I argue, that can be done easily by human beings. The purpose of the questionnaire survey is to learn how humans reason about geographical data. The challenge of geographic reasoning is that it must typically deal with incomplete information (Egenhofer and Mark, 1995). People can draw sufficiently precise conclusions, for example by completing missing information intelligently or by applying default rules, frequently based on common sense. In fact, Barkowsky and Freksa (1997) argue that people succeed in combining their general spatial knowledge with the contents of maps in such a way that an overall inference works even if the individual contributing pieces of knowledge appear deficient. We need to uncover how people conceptualise the physical environment, especially in relation to its use, how people reason and infer knowledge from a topographic map, and how we can best capture this knowledge. These concerns directly touch upon people's abilities and lives as seen earlier; they are not merely theoretical.

The survey presented here is not concerned with assessing spatial abilities of people (e.g. Smith and Mark, 2001; Mark *et al*., 1999). Neither does it address how cognitive mapping is developed and learned (e.g. Held and Rekosh, 1963; Orleans, 1973; Grittens, 1969), nor does it identify errors and distortions in spatial memory (e.g. Gehrke and Hommel, 1998; Lloyd and Heivly, 1987; Jahn *et al.,* 2005; Rothkegel *et al*., 1998). Instead, the interest lies with capturing spatial knowledge on land uses, to understand abilities and processes like grouping principles behind interpreting a topographic map according to function, and to identify people's internal conceptualisation of the spatial composition of land uses. Indeed, visual search processes used in map reading have been investigated (e.g. Board and Taylor, 1977; Barkowsky and Freksa, 1997) as well as the semantic meaning of land cover (e.g. Comber *et al.*, 2005a, 2005b, and 2005c). The aim here is to elicit ontologies from

human subjects to provide guidelines for developing links between high-level functional information and spatial data.

Central to this questionnaire survey is hence the interpretation of topographic maps and the associated reasoning process that operate on it (Kosslyn, 1978). Over the past, many cartographers embraced an experimental paradigm and studied the interaction between the map and map-reader (Freundschuh and Egenhofer, 1997). We can describe this interrelationship by the earlier introduced 'map in the head' metaphor, which, being inspected by the 'mind's eye', is functionally identical to a graphical map inspected by a 'physical eye' (Kuipers, 1982). This implies a direct relationship between the map's depicted reality and that of a map reader, as illustrated in figure 8 derived from Koláčný's (1969) communication model about cognitive aspects of cartography.
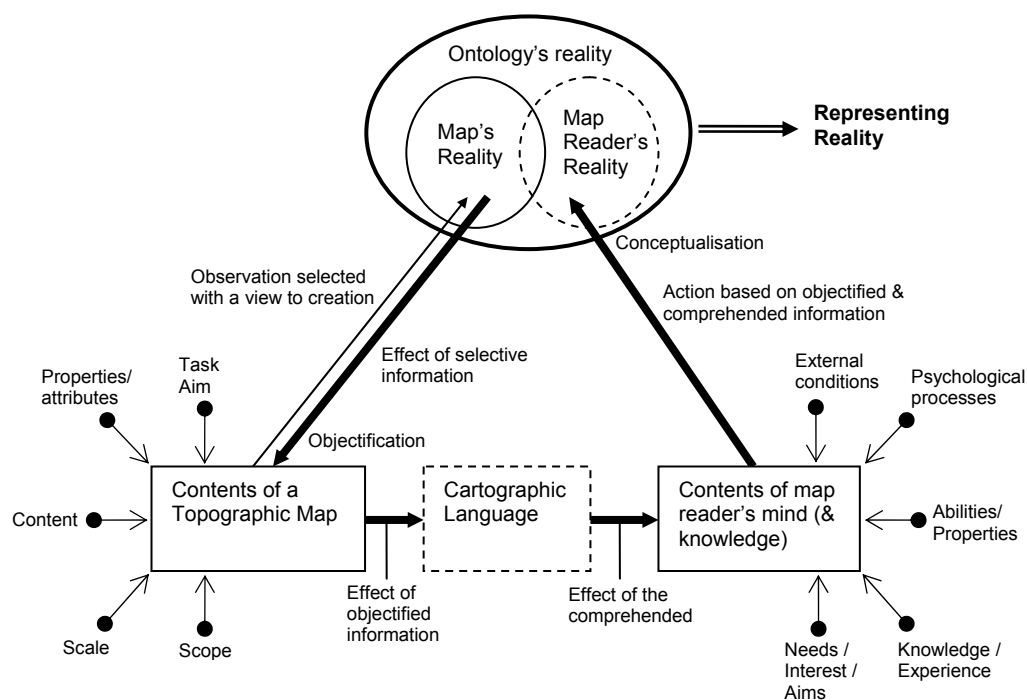
Figure 8 Mapping between a map's depicted reality and a map reader's reality

Originally, such a communication model was developed to systematise the process of cartographic communication by illustrating influencive factors between the cartographer's mind and the map reader's mind, to better understand resulting implications for map design and interpretation. Such a communication is not far off from what is to be achieved in this thesis. However, the central concern does not lie with the processes of communication, but with the map reader's expected view of a

real world representation, and that of the cartographer's map accurately depicted view. Both the map and the person inspecting the map carry a representation of reality determined by a variety of factors. By unifying these representations, we combine detailed spatial information from the map data with human-acceptable concepts that are intrinsically tied to their underlying geography, and thus to the data itself (Thomson and Béra, 2007a).

This thesis therefore relates human spatial perception of land use to the landscape characteristics. It accounts for the ways in which people represent and combine geographical information, how they recall it, and reason to derive new knowledge. As a result, the purpose of this undertaking is threefold:

1. To study a topographic map according to the processes that operate on it when it is being inspected by the map reader;

2. To study the nature of the input, or stimulus, perception and analytic processes and the nature of similarity judgement;

3. To study a person's conceptualisation according to the principles and structure of categorisation.

From this, we can induce relevant knowledge, capture and translate it into a machine-readable knowledge base (Thomson and Béra, 2007a). Later, an ontology can model this knowledge (see chapter 5) by explicitly stating how relevant concepts and their constituting objects relate to each other and manifest themselves in their physical existency in both reality and that of the representing geography.

Questionnaire design

Earlier in this chapter, we learned about spatial perception, cognition and categorisation of people. The related theories have important implications for the design of the questionnaire. The ability to think about one's own cognitive processes is fundamental to answering questions about the interpretation method, yet it remains a difficult task to make this knowledge explicit. Culture is both a ramification and manifestation of human cognitive abilities and, as such, cultural and social aspects will influence a respondent's way of thinking. It seems likely that answers will be a mere reflection of the spatial composition of land use types, because the physical environment and what it affords determines the input information for people and with that their knowledge. One may wonder why to perform a questionnaire survey in the

first place, if it only states the obvious. Nonetheless, its purpose is to capture knowledge, and a study of ways non-experts conceptualise a given domain of reality might help efforts to maximise future usability of the ontology, let alone through its empirical testing (Smith and Mark, 2001). Furthermore, if people's perception and representation of space differs individually due to the way they experience their environment, then it is important to generalise cognitive views across more than one person.

The first task of the questionnaire is to identify how humans interpret a topographic map for land use information. The map is studied in relation to the processes that operate on it whilst it is being read and interpreted. Interpretation requires the human ability to draw analogies from the familiar to the unknown. Because geographic kinds and concepts differ from everyday objects and kinds perceived by people, land use is not something people think about much in their daily lives. Nevertheless, Berendt and Barkowsky (1998) argue that operations performed on maps are routinely performed on internal representations making them natural and easy to accomplish. Of interest is how laws of perception influence and determine the map interpretation. In Barkowsky and Freksa (1997), a hierarchic order on different classes of aspects, or pieces of information, on a map is imposed for modelling interpretation processes. This technique reveals the depictional precedence of information used in the interpretation process. For example, existence and connectedness decrease in importance to distance and shape, which are only interpreted indirectly. Thus, the importance of map clues (i.e., shape, proximity, symmetry, contrast, etc.), context, scale, and perceptual criteria need to be evaluated as part of the interpretation decision. The questionnaire focuses on interpreting the geometry and configuration of mapped land cover parcels. Consequently, the topographic map needs to be stripped of all its additional colouring, cartographic text, scale and orientation information that is used to communicate the information in its entirety.

Existing nomenclatures of land use types have specific terminology for their categorisation, whereas spatial data only offer descriptions to space and its coordinate system. People do not refer to space but place, and thus substitute scientific geographical vocabulary with shared, everyday descriptions of place. To make data more accessible, it is important to learn which terms offer the most natural description

for people. This can be achieved by asking respondents to describe their interpreted areas with one word, not having them biased previously with land use terms. Furthermore, instead of making people just think plainly in terms of land use, the respondents should be asked how they would make use of an area and what the deliberate purpose is of an area. Thus, by having two maps interpreted according to use and purpose respectively, we will learn how people communicate about land uses, perhaps revealing a 'vernacular' geography for land use.

Another important aim is to study people's conceptualisation about land use, and generally the knowledge they have about its spatial organisation and relations. People reason about spatial problems with their mental representation composed of pieces of information, images or diagrams, and beliefs and emotions. Mental models can reveal themselves in two forms: as images or words. Language determines but also limits the way the world is perceived and conceptualised. The question is how people internally conceptualise a land use type's spatial composition. The questionnaire must offer a series of questions that capture the conceptualisation in written words and according to a structure that is similar to our innate structures of categorisation. As we learnt in section 3.1, categorisation is a means to achieve cognitive economy. It allows us to separate concepts into crisp categories. According to Rosch (1978), the vertical and horizontal dimension of our category system allows us to structure, organise and conceive our perceived environment. If we adopt a similar approach to this survey, then we can structure questions according to a horizontal dimension, where separate categories describe a land use spatially, and to a vertical dimension, where those categories are further described in detail according to a set of questions (Thomson and Béra, 2007a). This would look similar to the representation given in figure 9. The first question addresses concepts for a chosen land use type (i.e., the goal), describing it spatially according to other functions which make up that land use. This is then followed by questions addressing each concept's purpose, role and affordance (i.e., words defining its function), as well as its physical object of which the concept is made and its physical property and other relations. Consequently, taking a top-down approach from the general to the specific, the underlying land cover defines and represents high-level land use.

LAYERS	REPRESENTATION	GRANULARITY

GOAL — Knowledge — University — Generalisation

CONTENTS — Concepts — Campus   Refectory   Residential

Words — Leisure   Eating   Sleeping

Objects — Open space   Building

FACTS — Properties — Large, unevenly shaped   Small   Block-shaped

Relations — Adjacent   Disjoint

TOPOGRAPHY — Land cover — Natural   Man-made — Detail
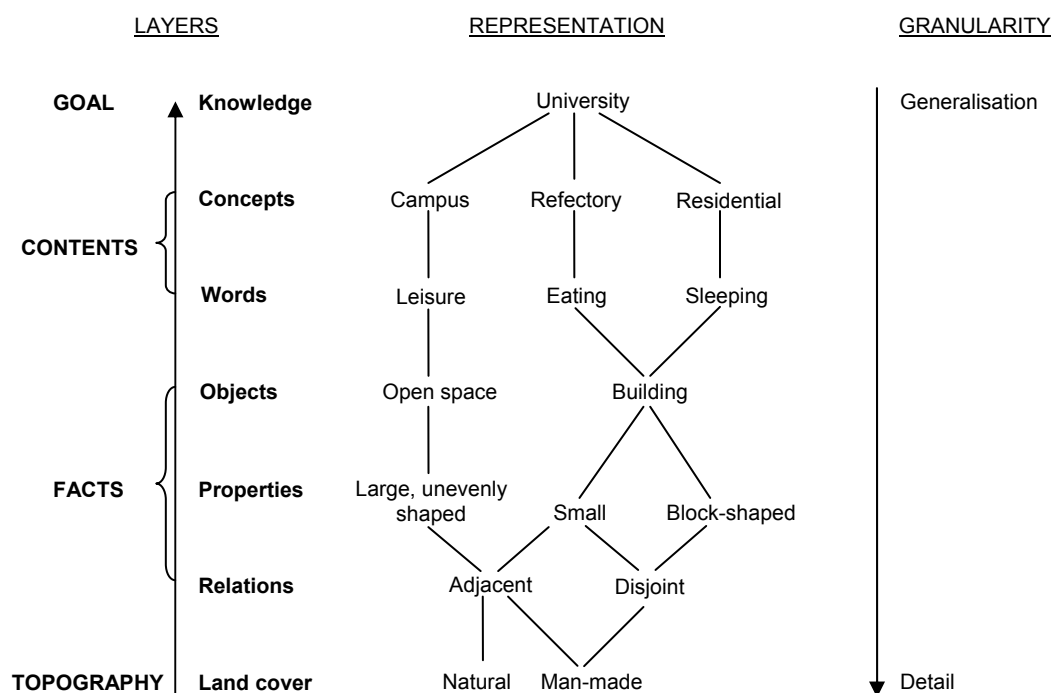
Figure 9 Top-down approach for capturing a person's inner conceptualisation

Altogether, the questionnaire consists of four tasks. Figure 10 illustrates a flow chart of the tasks highlighting procedures and interplay between them and their measures. The first task is to interpret two plain topographic maps according to use and purpose of their depicted areas. The expected output is information on relevant concepts to the ordinary map user, and if respondents are capable of interpreting 'successfully' a plain map for land use information. The second task evaluates the interpretation process of the first task by asking open-ended, closed and attitude questions. A person's approach to interpretation is measured considering difficulties, abilities and other factors. Important is the question whether the respondent is able to identify the location of the depicted area in the map. Ideally, this should remain unknown, so that it does not cause any bias in the interpretation. The chance of this happening has been reduced by stripping the topographic maps to their bare minimum and choosing a large scale. From this, it is expected to gain an insight into the reasoning processes behind interpretation and inferring knowledge. The third task captures a person's conceptualisation and knowledge of land use. The output will reveal how someone spatially conceptualises a given land use. The fourth and last task addresses the respondents' demographic information, which is required for analysing the questionnaire data. These tasks have been revised during pilot testing. The questionnaire can be found in appendix A.
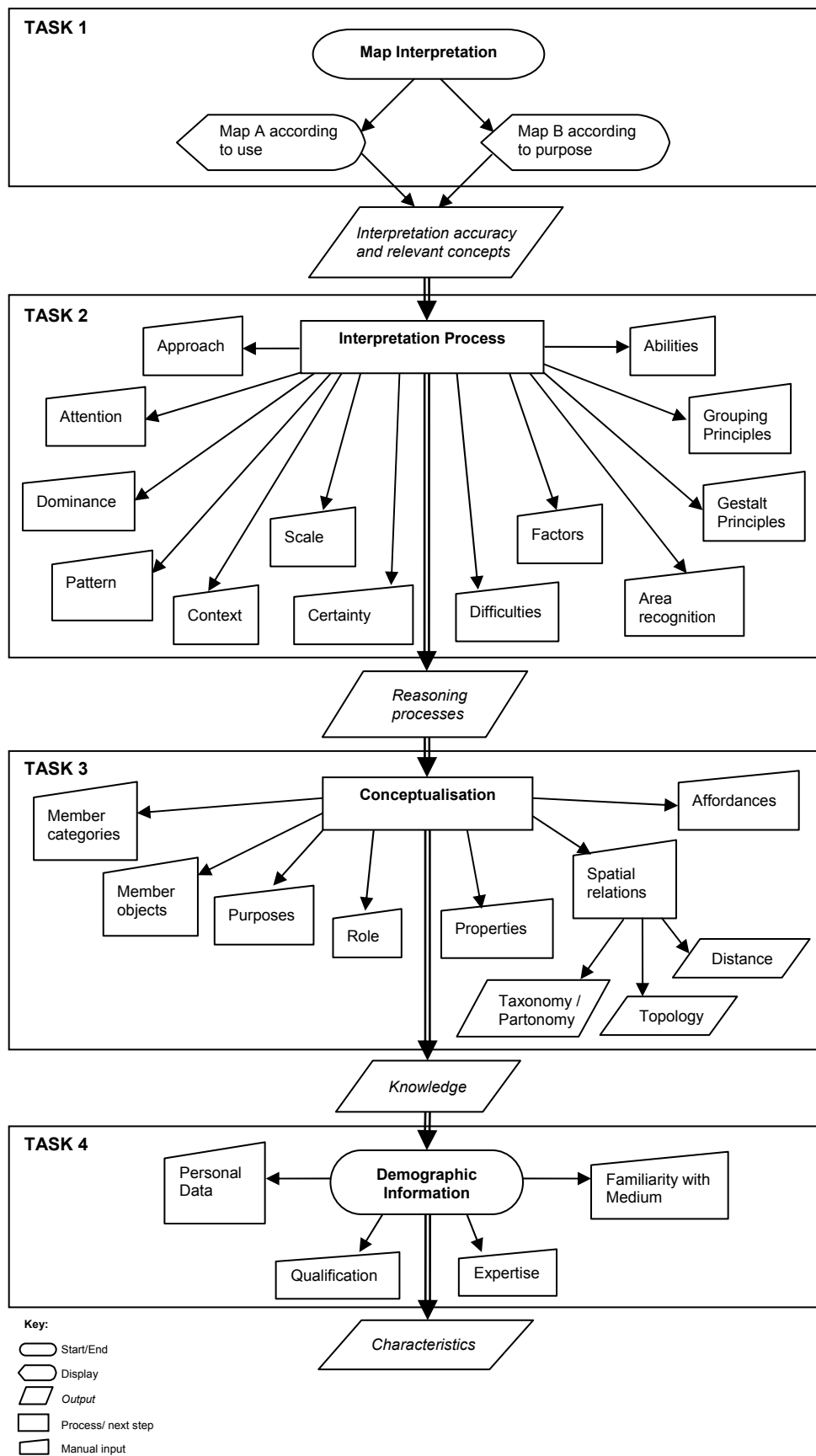
Figure 10 Questionnaire flow chart

Population and Sample

The survey aims at respondents that are familiar with geography and mapping. This deliberate choice of respondents ensures information-rich cases for in-depth study (Patton, 1990), and that respondents' skills and knowledge are sufficient for the tasks and questions within the survey. The sample therefore complies with the purposive sampling strategy for the explicit selection of respondents that will generate appropriate data, as opposed to statistical sampling strategies used in quantitative studies, which are more concerned with the representativeness of sample in relation to a total population (Pope and Mays, 1995). The pilot study involved a number of PhD students (N_p= 7) from UCL, and proved the usefulness and success of responses from this type of participants.

The sample size of such an explorative and descriptive study is dependent on the aims of the survey, availability of eligible respondents, and limitations of time and resources, as in contrast to probability samples where sample size can be calculated to minimise effect size and achieve precision (Green and Thorogood, 2004). The number of participants for the final survey aims at a small sample (a total of N_T= 18 participants) due to the labour intensive analysis of qualitative data and the difficulty in recruiting participants. Such a small sample is not representative and does not aim to test hypotheses. The survey therefore only provides indicative information on the respondent's interpretation and conceptualisation of land use. However, it is possible to carry out a wider research in future.

The majority of the sample is male (77.8%) and British (83.3%). The age distribution's majority is between the age of 18 and 30. A good mixture of participants took part with both varying educational levels and varying places of living and work including rural areas, towns and cities. Most respondents, more specifically half of them, are familiar with topographic data, whilst the rest is distributed among somewhat and a little of less familiar. The type of map data usage explains this high degree of familiarity. Maps are mainly used for personal and professional purposes. Ordnance Survey map products are used 77.8% of the time compared to other data such as street maps, digital maps, town plans, aerial photography, and terrain models. The frequency of map use has its majority between frequently (50%) and often (27.8%). More importantly, however, is that respondents did not recognise any of the locations depicted in the two maps to

avoid bias creeping into the map interpretation results. Despite the use of London data, and including participants from London, none of the participants identified the locations shown in the maps.

Collection of replies and analysis method

The data were collected by means of a self-administered, structured questionnaire. The method of distributing self-administered questionnaires allows participants to complete the questionnaire at their own time and convenience, besides being more cost effective than recruiting participants and performing interviews. The collected data were then categorised and coded to allow entering of the results into appropriate data files (Wall *et al.*, 2002). The assigned codes for the categories of each question, including codes for missing values and non-responsiveness, are compiled in appendix B. This includes the description and type of each question, its data type, measure level, variable names and labels, the value labels (i.e., code), any missing values, and the analysis method. Due to the qualitative nature of this study, we are dealing with string data and nominal and ordinal measurement levels. Especially with regard to the numerous open-ended questions, which provide free text descriptions, it is necessary to identify initial themes or concepts. By labelling and sorting data according to concepts or themes, we can detect emerging patterns and develop appropriate explanations. It requires reviewed decisions of where to be specific in terms of increasing the number of categories, or where to reduce similar answers into the same category, and hence loosing some of the richness of the data. Nonetheless, these iterative steps are typical analysis procedures in any qualitative analysis, as illustrated in figure 11 (Ritchie and Lewis, 2003). Such a methodical and standardised approach is crucial for ensuring a good qualitative analysis that is able to document its claim to reflect some of the truth of a phenomenon by reference to systematically gathered data (Pope *et al.*, 2000).

```
┌──────────────────────────┐  ┐
│ Seeking applications to  │  │
│ wider theory             │  │                          Iterative process
├──────────────────────────┤  │   EXPLANATORY            throughout analysis
│ Developing               │  │   ACCOUNTS                        ▲
│ explanations             │  │                                   │
│ (answering how and       │  │                                   │
│ why questions)           │  │                          Assigning data to refined
├──────────────────────────┤  │                          concepts to portray
│ Detecting patterns       │  │                          meaning
│ (associative analysis    │  │                                   ▲
│ and identification of    │  │                                   │
│ clustering)              │  │                                   │
├──────────────────────────┤  │                          Refining and distilling
│ Establishing typologies  │  │                          more abstract concepts
├──────────────────────────┤  │   DESCRIPTIVE                     ▲
│ Identifying elements     │  │   ACCOUNTS                        │
│ and dimension,           │  │                                   │
│ refining categories,     │  │                          Assigning data to
│ classifying data         │  │                          themes/concepts to
├──────────────────────────┤  │                          portray meaning
│ Summarising or           │  │                                   ▲
│ synthesising data        │  │                                   │
├──────────────────────────┤  │                                   │
│ Sorting data by theme    │  │                          Assigning meaning
│ or concept (in cross-    │  │                                   ▲
│ sectional analysis)      │  │                                   │
├──────────────────────────┤  │   DATA                            │
│ Labelling data by        │  │   MANAGEMENT             Generating themes and
│ concept or theme         │  │                          concepts
├──────────────────────────┤  │
│ Identifying initial      │  │
│ themes or concepts       │  │
├──────────────────────────┤  │
│ RAW DATA                 │  │
└──────────────────────────┘  ┘
```

Figure 11 The stages and processes involved in qualitative analysis

For analysis, the elicited data were entered into the statistical analysis software SPSS version 14.0. Due to the mixture of data types, a number of different statistical analysis options are available. For binary or yes/no answers, statistics such as Chi-squared, Spearmans, Wilcoxon, Mann Whitney, and Kruskal Wallis are useful. The rating or visual scale requires for example the T-test, Pearsons, Analysis of Variance (ANOVA), or cumulative frequencies and proportions. For open-ended replies, one can employ thematic content or discourse analysis, or also frequencies and proportions. The analysis of the survey results is mainly limited to running simple counts, frequencies, percentages, and row proportions due to the majority of categorical responses. Nevertheless, these simple statistics summarise the results, display the relative distribution of responses, and thereby identify emerging patterns and tendencies. Employing these standard analysis techniques ensures the results are valid and reliable for analysis.

Results and analysis

In the first task, participants were asked to carefully examine two topographic maps and to look for features, similarities and patterns that they may recognise. Then they should group those objects that they believed belong to the same land use category by circling or colouring in the area. Two scenarios were given for the interpretation. In the first map, respondents were asked to interpret the map according to how one can use the areas. In the second map, respondents were asked to think of urban planning, where everything is built for a specific purpose, and then to interpret the map according to the purpose of areas. In addition, they were asked to rate their confidence of the interpretation. Figure 12 illustrates one of the respondent's interpretations.
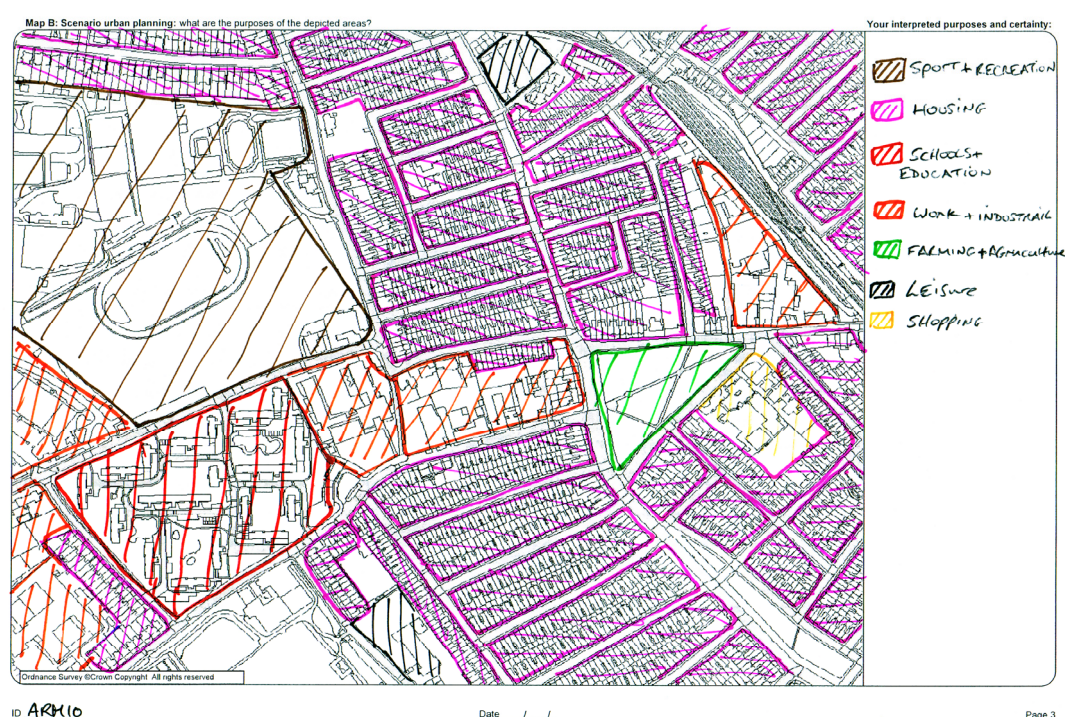
Figure 12 A respondent's land use interpretation

Table 2 and table 3 summarise the results in frequencies of the interpretation for maps A and B, respectively. The interpretations were compared against GeoInformation Group's Cities Revealed land use dataset. The land use types that each map contains are summarised in the top row of both tables. Each interpreted map is examined for not identified land use types, which ones were interpreted correctly and which ones were misclassified, as shown in the left column. In the case of correct interpretations, it was also examined whether respondents used different concepts or the same terminology as annotated in the GeoInformation Group land use dataset. The results from both maps

suggest similar outcomes across the given criteria. The majority of respondents did not identify land uses such as offices, car parks, institutional and communal buildings, storage and warehousing, industry, standing water, religious buildings, and indoor recreation – keeping in mind that participants were not given a list of land use types to look for, but to search by themselves for uses and purposes of the depicted areas. Similarities were also present in the misinterpretation of land use information in both maps. The majority of misclassifications include retailing, educational and institutional buildings, industry and offices. The most accurate and successful interpretations are residential areas, retailing (to a certain extent), railways and outdoor recreation. However, the focus slightly differed in the first map, where only eight participants identified residential areas compared to twice as many in the second map.

Table 2 Interpretation results for map A in frequencies

	Residential	Retailing	Offices	Railways	Car parks	Outdoor recreation	Educational buildings	Institutional buildings	Institutional & communal buildings	Storage & Ware-housing	Industry	Standing water	Religious buildings
Not identified	10	1	17	3	17	0	9	11	18	17	16	18	15
Interpreted correctly (same term used)	5	1	0	1	0	3	0	0	0	0	0	0	0
Interpreted correctly (different term use)*	3	5	0	13	0	14	3	0	0	0	0	0	2
Interpreted falsely	0	11	1	1	1	1	6	7	0	1	2	0	1
Total	18	18	18	18	18	18	18	18	18	18	18	18	18
*other terms:	• Cheap houses • Council estate • Housing • Housing estate • Nice houses • Residential area • Residential housing • Terraced houses	•Commercial •Hardware store •Local shops •Shopping •Shopping area •Shopping centre •Shops •Post office •Superstore/ retail park		• Railway line • Railway station • Train station • Transport • Transport links • Travel • Travelling • Underground station • Underground/ trains • Train	• Super-market car park	• Golf course • Leisure complex/centre • Park • Park area • Parkland • Playing field • Public park • Public open space • Recreation • Recreational • Soccer pitch • Sports area • Sports field • Sports pitch • Children park • Nature reserve	• Education • School • School playing field • School/ college • College • Secondary school	• Hospital • Health centre • Public building • Public house		• Storage/ park	• Factories • Industrial area • Small factories • Work outlets		• Church • Places of worship

Table 3 Interpretation results for map B in frequencies

	Residential	Retailing	Offices	Railways	Indoor recreation	Outdoor recreation	Educational buildings	Institutional buildings	Storage & Warehousing	Industry	Religious buildings
Not identified	2	9	15	10	17	0	13	13	17	11	17
Interpreted correctly (same term used)	8	0	0	3	0	11	0	0	0	3	0
Interpreted correctly (different term use)*	8	0	0	5	0	7	1	0	0	0	0
Interpreted falsely	0	9	3	0	1	0	4	5	1	4	1
Total	18	18	18	18	18	18	18	18	18	18	18
*other terms:	• Council/ government area • Flats/ housing estate • High rise flats • Houses • Housing • Housing estate • Structured housing • Tower-blocks/ mass housing	• Retail • Shopping • Shopping precinct • shops	• Purpose built offices • Business	• Railway line • Railway station • Railway station/ tracks • Train tracks • transport	• Indoor sport centre	• Athletic track • Entertainment • Fields/parks • Football stadium • Horse racing track • Leisure • Leisure centre • Leisure complex • Leisure facilities • Open spaces • Park • Park/ sports ground • Public open space • Public park • Recreation • Recreational facilities • Sport facilities • Sport/recreation • Sports area • stadium	• College/ university • School • School/ education	• Hospital • medical	• Warehouses	• Factories • Farming/ agriculture • Industrial • Work/ industrial	• Places of worship

Respondents' own rating of their interpretation confidence, according to being 'very confident', 'somewhat confident', and 'not confident', reflects closely the results of the interpretation accuracy. Figure 13 provides a graphic representation of the results from a cross-tabulation. Respondents were most confident in determining recreation, residential and railway, which were also the most successful interpretations. Likewise, respondents' least confidence reflects those categories that respondents mainly misclassified such as educational and institutional buildings, retail and industry. The results suggest the ambiguities and uncertainties involved in interpreting some land use categories.
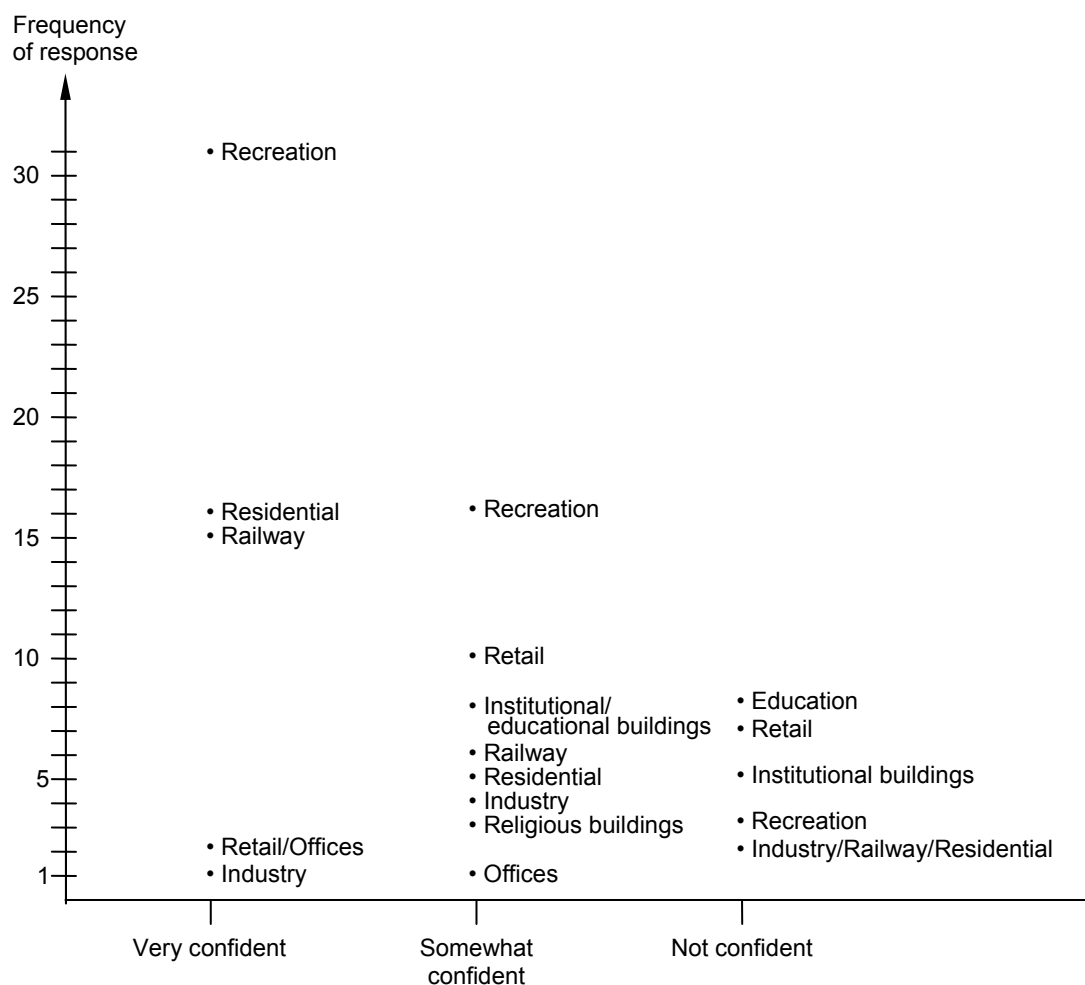
Figure 13 Respondents' judged confidence in their interpretation

For each land use category present within the maps, the majority of people referred to their interpretations in a number of different ways to the annotations used in the GeoInformation Group's dataset. This indicates that everyday people's use of terminology can be very different from professional geographic vocabulary in the land

use domain. As can be seen from table 2 and table 3, a remarkable 63.5% of correctly interpreted land use types were described with concepts different from GeoInformation Group's terminology. This not only indicates the generic richness and diversity of land use categories, but also that people communicate with much more common, but detailed concepts than those all-embracing professional concepts used in existing land use nomenclatures. For example, respondents rarely used umbrella terms such as education or outdoor recreation. Instead, respondents distinguished between schools, colleges and universities, or athletic track, football stadium, park, and playing field. This phenomenon can be observed for all of the GeoInformation Group's land use classification. People tend to speak in more specific terms by communicating as precisely as possible the meaning of the object they are referring to. In addition, these terms often reflect basic and simple concepts used in everyday language. For instance, rather than saying residential, respondents used the more common term housing or flats. As we learnt in section 2.3, professional users also require better-separated concepts because of implications for their applications. In fact, the results emphasize the general problem of semantic heterogeneity, and the need to include human acceptable concepts in spatial representations. This would not only ensure interoperability, but also improve spatial analysis and the general use of data.

The second task enquired about the clues, factors, and reasoning processes that contributed towards the interpretation. It is important to learn about these processes to identify useful parameters and key pieces of information relevant for automating the procedure of deriving functional information from topographic data. The first set of questions addressed the interpretation approach, attention and dominant objects. Respondents used a variety of approaches to interpret the maps:

- *"Shape"* (38.8%)
- *"Size"* (16.7%)
- *"Large objects first"* (16.7%)
- *"Similarities"* (16.7%)
- *"Searching for familiar areas"* (11.1%)
- *"Envisioned home town"* (11.1%)
- *"Familiarity with urban layouts"* (11.1%)
- *"Relationships between objects"* (11.1%)
- *"Knowledge from using maps before"* (11.1%)

Regarding the question what captured the respondent's attention first in both maps, similar answers were found despite the different areas shown. For example, in map A attention was mainly drawn to residential areas (16.7%), parks (44.4%), and other open spaces (16.7%). Some respondents also mentioned straight lines that were interpreted as railway lines. Equally, in map B residential areas (16.7%) and open spaces (11.1%) captured respondents' awareness. The oval track or sports ground was the most unique feature according to 66.7% of the respondents. Other clues about the interpretation include the following:

- 15 respondents (83.3%) made use of already interpreted areas for identifying further ones.
- 16 respondents (88.9%) believed there is a repeating pattern for each land use type.
- 10 respondents (55.6%) agreed that it would have helped if the map showed a bigger area.
- 6 respondents (33.3%) believed the varying scale of the two maps influenced their interpretation, while 9 (50.0%) disagreed, and 3 (16.7%) did not know of any difference.

Indeed, scale and how much of an area is shown in a map depends whether an overview with fewer detail or vice versa is more desirable. Those respondents who believed scale influenced their interpretation stated that this was because:

- *"at smaller scale, larger areas relate better on the map."*
- *"smaller scale more to interpret."*
- *"things look different."*
- *"at smaller scale buildings harder to interpret what they could be used for."*

In OS MasterMap, urban areas are represented at a detailed scale of 1:1250. The maps, however, were represented at a scale of 1:3000 and 1:4000, respectively, to increase context for the interpretation. The results suggest that respondents first classify areas they are familiar or confident with, and then move from there to the more uncertain areas. Despite some difficulties during the interpretation, most respondents believed that land use categories have a repeating pattern. The uniqueness and consistency of patterns is crucial for any attempt to identify land use types automatically according to configurations and context.

Uncertainties and difficulties in the interpretation reflect the land use categories that respondents mainly misinterpreted. Respondents felt that the biggest obstacle in the interpretation is that many areas could be interpreted as many different things, especially shops, amenities, and large areas or buildings. Other difficulties included the identification of public use buildings, generally identifying the detailed purpose of buildings, differentiating among residential and business, and industrial and commercial. One statement adequately summarises the findings: *"identifying anything other than open areas and residential areas was a guess"*. In particular, respondents seemed to agree with the following difficulties:

- 12 respondents (66.7%) agreed that difficulty was caused by the fuzziness of where one land use ends and the other starts.

- 15 respondents (83.3%) agreed that difficulties were caused by the misinterpretation of cartographic objects.

- All 18 respondents agreed that difficulty was caused by not being able to identify an object's meaning.

- 14 respondents (77.8%) believed that difficulties are caused by not being able to identify one area's meaning in relation to other areas.

- No respondents recognised the location of the areas depicted in the maps.

The delineation of which groups of objects belong to a land use category is challenging, because there can be multiple land uses for one object. This is not to say that the underlying reality is in some respect ultimately vague (Smith and Varzi, 2000), but that people's categorical scheme is an accreditation for a distinction between crisp and fuzzy denizens of reality. Delineating boundaries is a manifestation of people's ability for picking up patterns and grouping objects together according to similarities. Therefore, it is necessary to investigate those principles that contribute to people's perception to obtain clues about stimulus factors for the interpretation of land use types. Figure 16 shows the results for respondents' rated importance of Gestalt principles, addressing the relations among parts and wholes, spatial contiguity, proximity, similarity of shape and size, common fate, good continuation, common region, closure, and element connectedness. On a scale from 1 to 5, reflecting ordinal levels of 'not important', 'little important', 'somewhat important', 'a fair bit important' and 'very important', respondents were asked to rate which pieces of information or stimulus factors they thought were superior to others in the interpretation process. Initially, the results were

summarised in frequencies. Then, to derive a better picture of the trend in importance and unimportance of the factors, the ratings were combined into three reflective columns of 'little important or less', 'somewhat important', and 'a fair bit important or more'. From the combined frequencies, row proportions were calculated to draw a less cluttered bar chart that visualises respondents' answers. Figure 14 indicates that respondents felt that shape, size, similarity in arrangement and geometry, and simplification of identification were most important for successfully interpreting a topographic map. On the other hand, symmetry, context, orientation, and likelihood of correct interpretation were mostly rated as 'little important or less'. Despite a relative high rating of importance across all principles for the interpretation process, there are factors that are superior in importance than others.
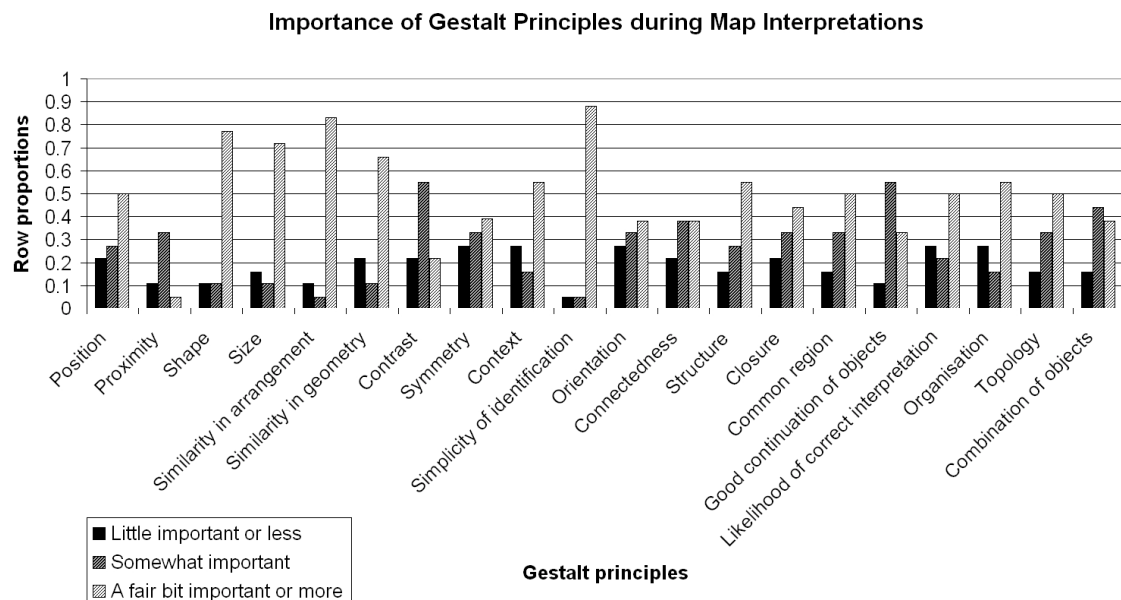
Figure 14 Rated importance of gestalt principles for the interpretation process

In addition, respondents were asked about the importance of principles for grouping objects together. The same scale was used as above with the calculated row proportions shown in figure 15. Similar to the results of the rated Gestalt principles, respondents felt that the most important factors were similarity in shape, size, and orientation. Proximity between objects, symmetry in arrangement, the relation among parts and wholes, and influence of one dominating feature within the group were somewhat important, whereas alignment of objects was rated as least important. Consequently, there is a clear tendency towards the importance of similarities rather than symmetries and alignments of objects. Additional comments by respondents showed that some felt grouping

principles are only relevant to residential areas. Others mentioned the distribution of roads as important, and that some types of building use are more likely to neighbour each other.

Importance of Grouping Principles during Map Interpretations
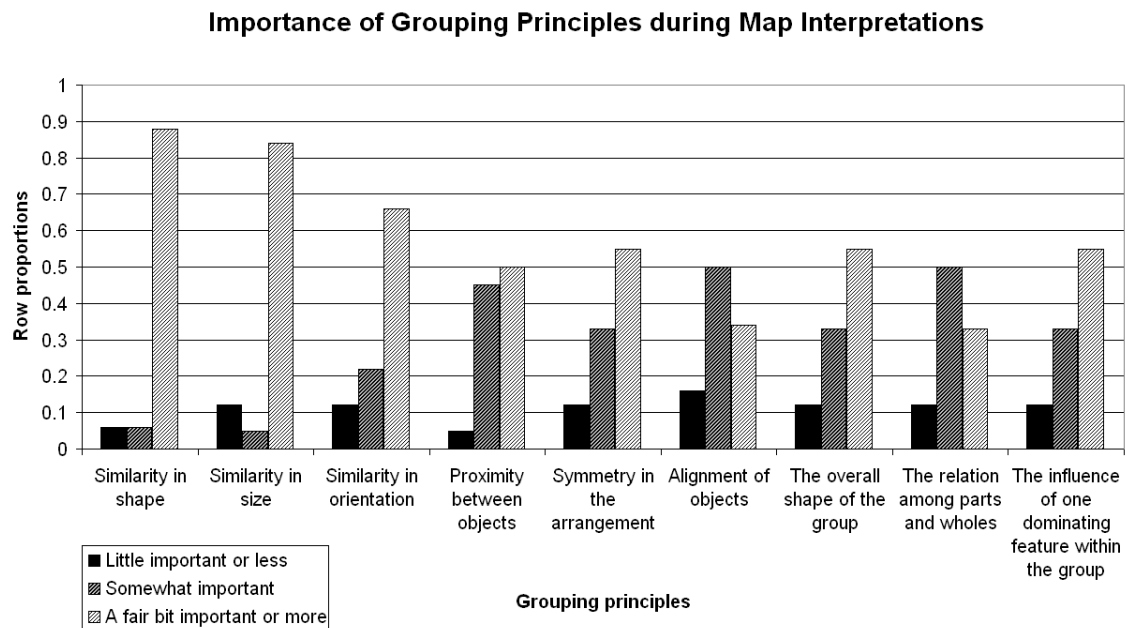
Figure 15 Rated importance of grouping principles

Important Abilities for the Map Interpretation
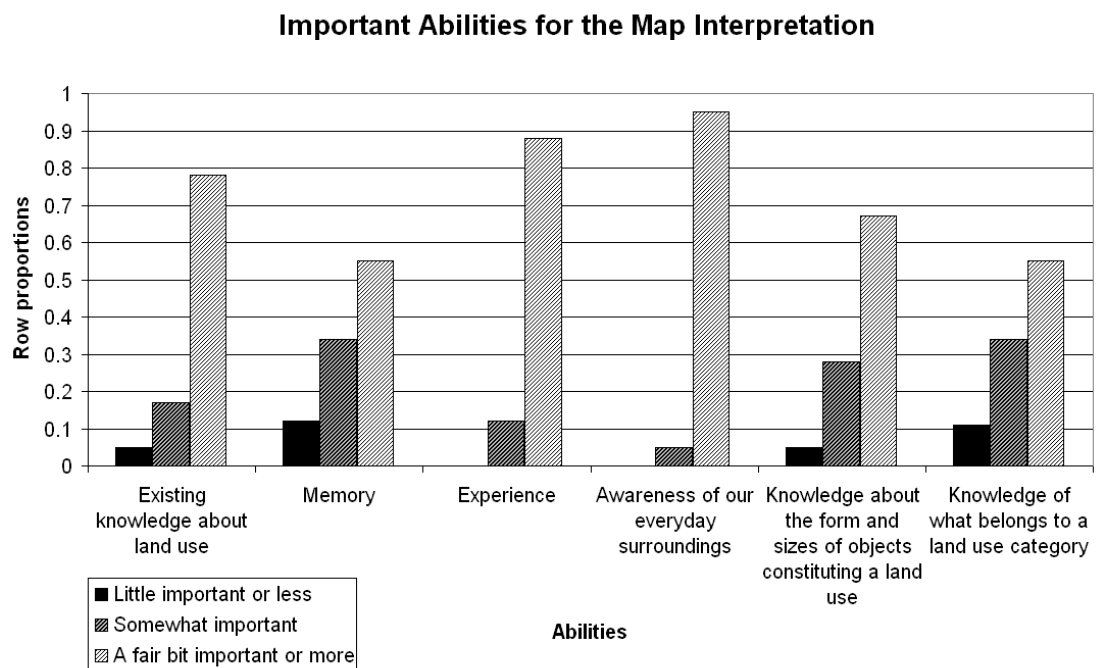
Figure 16 Importance of abilities for interpreting maps

Lastly, respondents rated the importance of a number of abilities to interpret the maps. Figure 16 shows the row proportions of the ratings. The results indicate that experience, awareness of our everyday surroundings, and knowledge about land use are most

important for deciphering land use information from a topographic map. In fact, all of the given categories received relative high ratings in importance, with only memory and knowledge of what belongs to a land use category being rated as 'little important or less'. These results confirm that interpretation is a knowledge-intensive task where experience is crucial for the inference through analogy, and that any expert system needs to consist of an exhaustive but precise knowledge base and a trained problem-solving engine.

In the third task of the questionnaire, respondents were asked to conceptualise a given land use with respect to its underlying landscape character. From a list of land use types respondents were asked to choose one and to think what constitutes a land use in the landscape. In the horizontal direction, separate categories define the chosen land use spatially, whereas in the vertical direction, a set of questions describe each member category in more detail. This task was specifically designed to acquire knowledge that will provide the skeletal foundation of categories, concepts, objects, relations and attributes for the conceptualisation and formalisation of an ontology in the land use domain. Due to the complexity of the task, a number of respondents either did not attempt this part at all or only parts of it. Fifteen out of the eighteen conceptualisations are useful for analysis. From the previous pilot study, where seven participants were asked to conceptualise six different land use types each, 42 conceptualisation were collected. This gives a total of 57 conceptualisations across the land use types industrial area, educational institution, hospital, recreational area, train station, and residential area, leading to a total of 285 member categories, and an overall total of 2565 concepts.

Table 4 summarises the frequencies of the land use type's associated member categories. The member categories are a reflection of the spatial footprint of land uses, which consist of other geographical objects and functions. The spatial component constrains the number of member categories that suit a specific land use type, therefore resulting often in similar conceptualisations. As table 4 illustrates, for each land use type there are a number of frequently occurring member categories. For example, land use education consists of the frequent member categories car park, classroom, sports field, gym, playground, school building, and canteen. Industry on the other hand consists of factory, car park, office, warehouse, roads, park, and shops. Some member categories are typical for their land use category, whereas others are common across

different land use types (e.g. car park). For each of the land use categories there are a number of no responses, which means that participants did not always think of all possible land use categories in the space provided. Furthermore, the number of examples given by respondents per land use category indicates both the familiarity of the category itself and the richness and diversity of familiar category members (Smith and Mark, 2001).

Table 4 Member categories of land uses (frequency)

Land use	Education	Industry	Recreation	Hospital	Train station	Residential
Member categories	Car park (9) Classroom (5) Sport field (5) Gym (4) Play ground (4) School building (4) Canteen (3)	Factory (6) Car park (5) Office (4) Warehouse (3) Roads (3) Park (2) Shops (2)	Pathways (3) Park (2) Playing field (2) Small building (2) Trees (2) Lake/pond (2) Fence/hedge (2)	Ward (7) Car park (6) Surgery (3) Waiting room (2) A&E (2) Reception (2) Subordinate department building (2)	Platform (6) Trains (4) Ticket office (4) Shops (4) Tracks (4) Station building (3) Car park (3)	House (8) Park (6) Garden (3) Roads (3) School (3) Shops (2) Block of flats (1)
No response	6	4	7	4	6	8
Total	66	40	49	45	45	40

Humans have a broad understanding of meanings according to contexts that allow them to overcome semantic heterogeneity, which is the main constraint on data interoperability. This is important because the term function is not a mere synonym for land use. It refers to an object's use, purpose, affordance and role, each of which it can have more than one. Functions are abilities that the object supports because of its deliberate design or purpose, but the object's affordances do not necessarily relate to these planned functions, which are realised through its specific use. The way people apprehend function, and how they discern the purpose, role and affordance of an object can be analysed and quantified by assessing the semantic relatedness among the concepts' terms. The majority of respondents found the differentiation between use, purpose, role, and affordance difficult. In the map interpretation task, participants were asked to look for uses and purposes in the two maps respectively. Although this initiated their own use of concepts for describing their interpretation, the differentiation does not appear to have much an impact on the interpretation. This is also evident in the third

task, where respondents defined the purpose, role and affordance of their stated member categories.

To learn how much similarity there is between the stated purpose, role, and affordance of each member category, we can calculate the semantic relatedness between the respective terms[1]. For this purpose, each pair of terms for purpose/role, role/affordance, and affordance/purpose was entered into the WordNet::Similarity web interface[2], which quantifies the degree to which two word senses are related. It provides six measures of similarity and three measures of relatedness, all of which are based on the lexical database WordNet (Pederson *et al.*, 2004). For this analysis, the simple node-counting measure path length is used to calculate semantic relatedness. The relatedness score is inversely proportional to the number of nodes along the shortest path between the synsets. The shortest possible path occurs when the two synsets are the same, in which case the length is one. Therefore, the measure can score between zero and one, the latter indicating high semantic relatedness. From the 216 terms that were given by respondents (excluding no responses), 45.8% of the terms were the same across all three, thus reaching a score of 1. However, of the 54.2% of the terms that differed, only a few cases had the highest score of 1 in semantic relatedness. This includes make/work, get/go, and drive/movement. In all other cases of differing terms, the highest similarity achieved is 0.5, going as low as to 0.0588. The average of the three comparisons indicates that purpose versus role has the highest semantic relatedness of 0.6788, followed by role versus affordance with 0.5860, and then by affordance versus purpose with 0.5314. However, these averages are misleading because they include the terms that did not differ across the three concepts purpose, role and affordance. Therefore, the cases that consist of the exact same terms are excluded from the analysis to get a better picture of the semantic relatedness between terms that differed.

A sample of 19 cases was further investigated. Table 5 summarises the results showing the member categories to which the comparison of purpose/role, role/affordance, and affordance/purpose relate, and their attained score of the semantic relatedness measure. The new calculated average indicates that terms for role and affordance are more related than terms compared in the other two cases. In figure 17, the line representing role

---

[1] For analysis purposes, only the results from the final survey are taken.
[2] Available from http://marimba.d.umn.edu/cgi-bin/similarity/similarity.cgi

versus affordance lies above the other two. However, overall the semantic relatedness among terms that differed between purpose, role, and affordance is low. Although these terms do not relate much semantically, it seems that respondents generally do not differentiate much between purpose, role and affordance as indicated by the high number of the same terms used across all three concepts.

Table 5 Semantic relatedness among terms that differed

Category	purpose	role	score	role	affordance	score	affordance	purpose	score
School	teach	learn	0.1429	learn	educate	0.2000	educate	teach	0.1110
Factory	production	make	0.1250	make	work	1.0000	work	production	0.5000
Hospital	operate	treat	0.1667	treat	receive	0.1667	receive	operate	0.2000
Industrial - open space	recreation	relaxation	0.2000	relaxation	smoking	0.1429	smoking	recreation	0.0740
Train station building	transport	commercial	0.0833	commercial	travel	0.0909	travel	transport	0.2500
Rail lines	network	transport	0.2500	transport	travel	0.2500	travel	network	0.1667
Roads	movement	car	0.1250	car	drop-off	0.1110	drop-off	movement	0.2500
Car park	revenue	parking	0.0909	parking	pay	0.1667	pay	revenue	0.0667
School - playing field	recreation	education	0.3333	education	club	0.1000	club	recreation	0.0909
play ground	play	break	0.3333	break	exercise	0.3330	exercise	play	0.5000
teaching facility	educating	teaching	0.5000	teaching	shelter	0.1250	shelter	educating	0.1250
gym	sport	meet	0.2500	meet	disco	0.2000	disco	sport	0.1667
play ground	play	relaxation	0.2000	relaxation	leisure	0.5000	leisure	play	0.1667
Residential - garden	recreation	entrance	0.1667	entrance	walking	0.2000	walking	recreation	0.1250
Recreation	sport	gathering	0.2500	gathering	meeting	1.0000	meeting	sport	0.2500
Train station	ticket	travel	0.1667	travel	paying	0.2500	paying	ticket	0.2500
Car park	storage	parking	0.0833	parking	driving	0.2500	driving	storage	0.1667
School - grass area	football	play	0.2000	play	exercise	0.5000	exercise	football	0.1667
Shop	selling	buying	0.3333	buying	display	0.2000	display	selling	0.2000
		Average:	**0.2105**		**Average:**	**0.3045**		**Average:**	**0.2014**
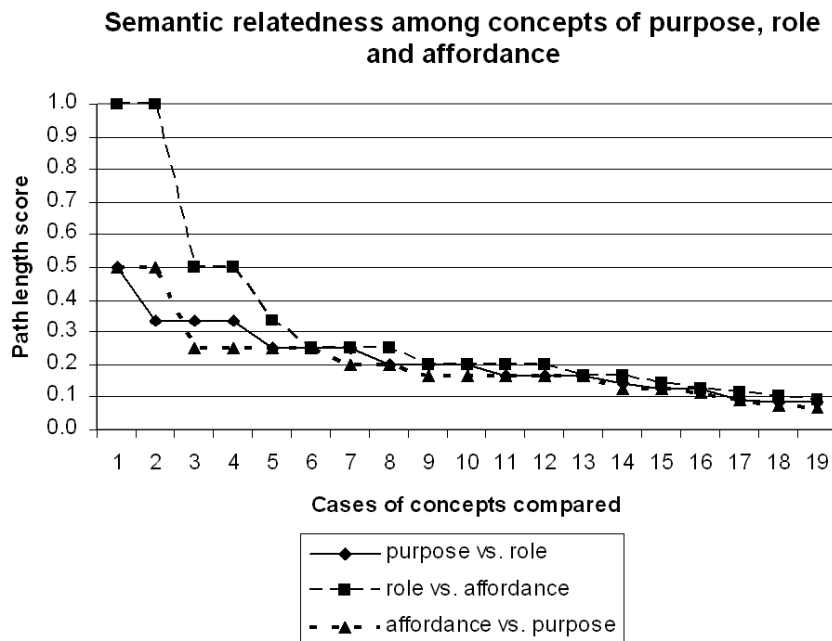
Figure 17 Semantic relatedness among terms that differed

Discussion

The subject under investigation needs to be treated carefully, as mental representations and individual knowledge of the respondents are subject to accidental properties, which should not be confused with the real properties of the phenomenon being studied. The other issue is order effect, which this survey does not account for. One task may directly affect the way that respondents approach the next one, which ideally should be eliminated. However, this is questionable, and if information from previous tasks influences subsequent ones, it will be expected to have a minimal effect.

Overall, from the results of the map interpretation task, together with the findings from the pilot study, we can conclude that humans can interpret land use information from plain topographic maps to varying degrees of success. Although the interpretation results may improve with more clues in the data such as colouring or cartographic text, the significance of this task is to learn more about humans' ability to infer land use information solely from its spatial configuration and context. Furthermore, it seems valid to postulate that some land use categories have unique and repeating patterns, like residential and recreational areas. Spatial configuration is an important characteristic of a land use's spatial footprint, whose extent is largely determined by grouping objects together according to a set of Gestalt principles. Some features and patterns dominate the depicted areas, hence capturing the participants' attention first. Respondents systematically searched the maps for what they believed made up a specific land use type. By considering shapes and sizes, and buildings versus open spaces, respondents thus identified the most familiar land uses followed by the less confident ones. Therefore, the key lies in the use of spatial relations, contextual and perceptional information to establish a complete definition of land use types. In fact, often a purely parameter based approached derived from data mining and pattern recognition techniques will not suffice in the interpretation of specific functions (e.g. Hussain *et al.*, 2007).

The survey emphasises the need for experience and knowledge about the domain of interest. With knowledge being at the heart of any implemented knowledge-based system, this stresses the importance of capturing relevant knowledge and translating experience into trained mechanisms. The questionnaire survey offers a great means to source knowledge from a larger set of people – although, finding volunteers can be

challenging at times. High rates of non-responsiveness indicate that extracting knowledge about land use is not as straightforward as asking people about their general opinions and typical habits.

Generally, qualitative research has been criticised for lacking scientific rigour, reproducibility, and generalisability (Mays and Pope, 1995). Indeed, the survey's major weaknesses relate to its small sample, limited analysis options, and the data's context specificity, which means that the results cannot be generalised to other contexts. On the other hand, complex data can only be acquired through small samples, as otherwise the workload becomes too labour-intensive and unmanageable. In addition, the detailed and context-rich data are particularly valuable for ontology engineering. The questionnaire survey therefore contributes towards techniques for knowledge acquisition and testing the use of non-experts as a source of knowledge. In fact, questionnaires for ontology elicitation have not been much studied or employed except for some experimental designs as for example in Agarwal (2004).

There are many different ways of capturing a person's knowledge depending much on the type of domain. This may involve in-depth interviews (Ritchie and Lewis, 2003), focus groups (Kitzinger, 1995; Green and Thorogood, 2004), or written questionnaires (Wall *et al.*, 2002). However, for specific knowledge acquisition, a purposive sample is crucial to ensure familiarity with the domain. For example, the results seem to fortify the findings in Smith and Mark (2001) where the number of examples given by respondents per land use category reflects some combination of the familiarity of the category itself and the richness and diversity of familiar category members. The study shows that people think in many different ways resulting in different conceptualisations and concepts. These may only differ slightly, yet they accentuate the problem of semantic heterogeneity. The conceptualisations are replicas of our spatial environment, because of the domain's spatial characteristics and strong influences from the natural world. Nevertheless, the number of concepts that respondents used for describing land uses and purposes in the map interpretation indicate that people communicate in terms that are most familiar to them. Consequently, the identification of human acceptable and familiar concepts should be the first step towards establishing more realistic models and representations of space.

The survey highlights this current gap between existing spatial models and the way humans interact with and conceptualise space. This thesis forms a bridge over this gap by relating human spatial perception of land use to the landscape characteristics. The identified concepts will be useful for describing land use types and will contribute towards a domain ontology of geographical knowledge, serving both the need for interoperability and information retrieval in future. This, however, may not be straightforward when dealing with conceptualisations from different countries and cultures, where the spatial environment, customs, mentality, etc., differ from our own ones. Although this assumption has not been tested here, it would be interesting for a further study to evaluate variations among socio-economic groups across different cultures. Real culture-induced differences can only be analysed by comparing findings to other groups abroad in relation to their appropriate topographic datasets.

Conclusions

The motivation for this survey has been the quest to learn from our own reasoning processes and abilities when interpreting land use information from topographic maps. Capturing the essence of people's knowledge is important for expanding sources of expert knowledge and deriving human-acceptable concepts for ontology engineering. Despite difficulties in making this knowledge explicit, the cognitive and qualitative approach to relating land use to the landscape character proves useful. Qualitative research is firmly established within social sciences. We have been stepping on *terrae incognitae* by pondering over disciplines such as cognitive science, psychology, philosophy, linguistics, computer science, and anthropology. The combination of cognitive sciences with information theory and GIScience helps to understand human interaction with and conceptualisation of functional space. As people's perception and conceptions not only vary among each other but also with the established spatial representations, human-based concepts need to be reflected in a more timely, realistic and acceptable manner.

Humans use multiple mental models of the world to reason efficiently at different levels of abstraction. Current geographic information systems normally use only a single model or representation of the world. Hence, the use of the system is limited. GIS needs to support multiple representations at different levels of abstraction, so that adequate

levels of abstraction can be found for a large range of scales. For example, Mennis *et al.* (2000) argue, if the cognitive representations and acquisition of geographic knowledge is divided between sensory information and derived knowledge, so too can the geographic database representation be divided between spatiotemporal data and higher-level geographic knowledge derived from that data. Consequently, this field has developed with two complementary problems: The theoretical problem requires a better understanding of how humans perceive the world and acquire higher-level spatial concepts. The practical problem addresses how to make computers interpret spatial information (Peuquet, 1988). This thesis is not just addressing the mere integration of semantics into database representations. It looks at cognitive principles to enrich spatial representations automatically.

In the quest to relate land use to the landscape character, we are addressing both the geometric structures of space models and their corresponding human conceptualisations. As Peuquet (1988, p.377) identified: "*it seems that research regarding cartographic representation has historically progressed along the two separate tracks with no significant integration. These two tracks coincide with the dual aspects of the map* […] *as geometric structures and maps as images.*" This survey focused on the latter, on the 'map in the head' metaphor and people's conceptualisations. Perhaps this is the first step from technical feasibility to the social acceptable, where we incorporate elusive measures such as human values, attitudes, beliefs, judgement, trust and understanding. On the one hand, this will ensure the development of a good ontology because of the empirical testing of non-expert knowledge, which should help to maximise the usability of the ontology and corresponding information systems (Smith and Mark, 2001). On the other hand, the use of non-experts ignites potential infiltration of erroneous beliefs within the captured knowledge (Smith, 2004). Even if we put confidence in some machine being able to make interpretations that actually correspond to some meaningful state of the world, we can only derive true statements from other true statements. All knowledge-based approaches remain limited insofar that these systems narrowly focus on specific domains of knowledge and cannot venture beyond them. As a result, their performance is always based on the accumulation of a substantial body of task-specific knowledge, motivated often by a combination of science and application on real-world tasks. Their

success is determined at least in part by accomplishing a useful level of performance on that task.

Questionnaires potentially offer a useful instrument for capturing knowledge despite existing pitfalls commonly known in questionnaire design (Boynton and Greenhalgh, 2004; Wall *et al.*, 2002). Considerations need to be taken towards the nature of knowledge, categorisation and conceptualisation in terms of cultural and linguistic constraints, and general difficulties in acquiring a person's believed knowledge. Therefore, the results are not to be understood in relation to a socio-economic group's representativeness, but only in relation to the derived knowledge. This knowledge, the key processes and factors involved during map interpretation, and how people construct representation of land use at different semantic levels, needs to be formalised into mechanised ways. The categorisation of functional roles and geographical entities can be easily represented by an ontology, as we will learn in the next chapters. Cognitive aspects of spatial relations can be formalised among concepts. Existing computational models inform the design and implementation of a computerized system that will be able to use these models for reasoning about functional information. The issues of knowledge representation and understanding of the spatial cognitive processes involved, the examination of respondents' views, and the theoretical aspects of cognitive science put this thesis in theoretical as well as applied contexts. Inference especially from a topographic database is not easy and the success of a knowledge system for reasoning about functional information in topographic data has yet to be proven. Even though, it is anticipated that domain specific knowledge possibly holds the key to enhancing cartographic data, and with that Ed Feigenbaum's gnomic dictum comes to the fore: *"Knowledge is power"* – indeed a true statement.

Chapter 4

The Ontology-Driven Approach to Enriching Spatial Databases

"To the extent that rational thought corresponds to the rules of logic, a machine can be built that carries out rational thought. [...] To the extent that thought consists of applying any set of well-specified rules, a machine can be built that, in some sense, thinks. ... A single machine can be programmed to do anything that any set of rules can do."

–Pinker (1997, pp.67-68)

The human mind is a complex but ingenious piece of natural engineering, which makes any so far attempted computational versions look pallid. Indeed, one major critique has been that computers are serial, doing one thing at a time, while brains are parallel, doing millions of things at once. Whereas computers outperform humans when it comes to doing repetitive tasks, humans are much more efficient in reasoning. Nevertheless, a computational theory of the mind states that all beliefs and desires are information, incarnated as configurations of symbols, which symbolise things in the world and are triggered accordingly (Pinker, 1997). Thus, we can speak of two types of representations, that of human versus computer knowledge representations. Although varied information is easily integrated and reconciled by human beings when needed and required knowledge is extracted, how does this work in a knowledge-based system? Take for example the World Wide Web with millions of web pages whose information volume rapidly grows making it increasingly difficult to find, organise, access and maintain information. To overcome these limitations, Tim Berners-Lee envisions the Semantic Web where meta-information annotates and defines the contents of a web page in a machine processable way (Davies *et al.,* 2003). The aim is to build knowledge and understanding from raw data, hence linking information in a more meaningful way. For example, when you enter a search term, instead of a search engine retrieving results that are varied in relevance, the semantic web 'knows', crudely speaking, which information to look for. From an Artificial Intelligence (AI) perspective therefore, knowledge representation refers to the encoding of knowledge in a form that can be

processed by a computer to derive consequences. Ergo, representations must be made of a formal, logical form, such as sentential logics, semantic networks, frame-based or object-oriented programming languages.

The analysis of mental representation, deductive reasoning, philosophy of language, and philosophical logic, as we partially touched upon in the previous chapter, have all contributed to building computational models of cognition through encoding information into knowledge representation (Wilson and Keil, 1999). Quite rightly so, as these computational models provide information that needs to be understood by their users and interpreted in the way the providers intended it (Kuhn, 2004a). We learnt that there is no single view on the world, but that there is a common basis of understanding through shared languages. Therefore, terms from natural language can be assumed to be a shared vocabulary relying on a common understanding of concepts with only little variety. The way the world is organised constitutes this common understanding. Conceptualisation is what we conceive it to be, this way or that way, and not some other way. It is a way of thinking about part of the world to which a limited number of persons commit at a time (Stuckenschmidt and van Harmelen, 2005). Conceptualisation thus represents ways in which we humans understand the world. For example, two different terms can be used to describe the same thing, as in English 'apple' or in German 'Apfel', but both share the same conceptualisation, a common understanding. In a GIS, conceptualisation naturally relates to some abstract description of geographical phenomena and concepts, such as building, land parcel, road, etc. (Smith and Mark, 1998). However, with the use of a shared terminology according to a specific conceptualisation of the world much information remains implicit. A vocabulary of terms is needed with some specification of their meaning.

Ontologies have set out to overcome the problem of hidden knowledge by making the conceptualisation of a domain explicit. An ontology is used to make assumptions about the meaning of a term, and as such ontology plays an integral part of knowledge representation. Knowledge representation is rooted in both epistemology, that is, the nature and sources of knowledge, and ontology, the study of the organisation and nature of the world independently of the form of our knowledge about it. The usual logical interpretation of epistemology is that knowledge consists of propositions whose formal structure and inferential aspects are the source of new knowledge. The notion of

ontology, on the other side, has two mainstream views (Agarwal, 2005): The original, philosophical view refers to Ontology (conventionally written with a capital letter) as a particular system of categories accounting for a certain vision of the world. In this sense, Ontology is the study of existence. In AI however, ontology refers to an engineering artefact, constituted by a specific vocabulary used to describe a certain reality, as well as a set of explicit assumptions regarding the intended meaning of the vocabulary words. A formal ontology is therefore the systematic and axiomatic development of the logic of all forms and modes of being (Guarino, 1995). Its specification can range from a simple catalogue or glossary of terms to the axiomatic theory of terms. Often a formal ontology is implemented in a knowledge base to facilitate intelligent reasoning, information retrieval or semantic annotation of data. AI researchers have been mainly interested in the nature of reasoning rather than in the nature of the real world. Reasons for the lack of interest towards ontology in AI research lies in the fact that problems like ontology and conceptual modelling need to be studied under a highly interdisciplinary perspective. The term ontology therefore tends to be used more to denote the content of a particular top-level knowledge base rather than to indicate a scientific discipline or a methodology.

Ontologies are useful for many different applications, but they all share the same idea. Ontologies help to reach a common understanding of a particular domain by identifying categories, concepts, relations and rules. These define and conceptualise the knowledge in a domain to model and provide a standardised vocabulary. The resulting specification of the meaning of this vocabulary of terms indicates how concepts are interrelated, and collectively impose a structure on the domain constraining the possible interpretations of these terms (Agarwal, 2005). Ontology therefore offers a means to improve communication between either humans or computers. Keita *et al.* (2004), for instance, summarise the use of ontologies as 'communication between humans and machines', 'structuring and organising knowledge', and 'reasoning by inference, particularly in very large databases'. Communication demands an explanation of the terminology used. System engineering benefits from a precise description of information and systems, which helps to identify requirements as well as inconsistencies in a chosen design. As we learnt in chapter 2, the reuse of existing software relies on specifying knowledge about existing components that can match the requirements of a given task. The ability to exchange information at run time, also known as interoperability, poses the same

demands as communication but among computers. Using ontology for the description of available information as well as for query formulation serves as a common basis for matching queries against potential results on a semantic level, thus facilitating information retrieval (Stuckenschmidt and van Harmelen, 2005).

The primary concern of knowledge engineering is to model systems in the world (Guarino, 1995). In GIS, the concern focuses on modelling spatial systems and geographical phenomena. GIS always had some sort of specification of the semantics of its represented features, take for example feature object catalogues or other land use and land cover nomenclatures. However, now that both data and methods may be retrieved and combined in an ad hoc way from anywhere in the world, these locally held specifications differ from other sources, are usually not machine-readable, and thus prohibit sharing with other systems (Kuhn, 2005). These concerns led to the emerging UCGIS research theme 'Ontological foundations for Geographic Information Science' (Mark *et al.*, 2004), which declares that research priority should focus on the semantics of geospatial information, in particular on the relations between human minds, information systems, and the geospatial world beyond. This thesis is concerned with the land use domain and its physical manifestation in topographic space. So far, we have studied and elicited the relation between human conceptualisations of land use phenomena and their real world representations. The goal is to ground a land use conceptualisation, for example residential area, in the topographic data representation, and to use ontology for an automated, semantic annotation of the data with functional information. Therefore, three research themes identified in Mark *et al.* (2004) are particularly relevant to this thesis:

1. The clarification of the relations between human knowledge, beliefs and representations on the one hand, the models and representations embedded in our data systems on the other hand, and the real world of objects beyond.

2. Research in eliciting geo-ontologies from human subjects (both experts and non-experts) using standard psychological methods.

3. Research in methods and tools for describing, accessing, and inferring semantic information from existing geo-spatial data.

The third research theme addresses the use of ontologies for information retrieval from spatial databases. In the subsequent chapters, we will obtain a visual demonstration of the richness of ontologies. First, however, we will learn how ontologies provide the

necessary means for translating our gathered knowledge about land use conceptualisations into a machine-readable format. The next section exemplifies how formal ontologies, specified in a logical theory, benefit spatial data. It will become clear why notions like semantics and ontologies have received so much attention within and outside geospatial information communities, but it will also highlight what ontologies can and cannot do. Then, the thesis applies knowledge engineering techniques and ontologies to the particular problem of inferring functional information from topographic data. Consequently, this chapter defines the solution, how this is a new step and if it is an isolated effort. Because ontology refers to the logical theory as applied in AI in the context of this thesis, I adopt the convention of writing ontology with a lower-case. Further, because ontology is not representative of a singular overriding truth, as in its philosophical sense, the use of plural ontologies is relevant and indicates multiple systems of conceptualisations.

4.1 How spatial data could benefit from ontologies

Although there are differences within ontologies, general agreement exists between ontologies on many issues: There are *objects* in the world that exist in various *relations* with each other. Objects have *properties* or *attributes* that can take *values*. Properties and relations can change over *time*. Objects can have *parts*. The world and its objects can be in different *states*. There are *processes* in which objects participate and that occur over time. *Events* occur at different time instants and can cause other events or states as *effects*. The representational repertoire of objects, relations, states, events, and processes does not say anything about which classes of these entities exist. The modeller of the domain makes these commitments (Chandrasekaran *et al.*, 1999).

In the geospatial domain, researchers have often asked what makes spatial special (Anselin, 1989; Egenhofer, 1993). Smith and Mark (2001) claim that one of the most important characteristics of the geographical domain is the way in which geographical objects are not merely located in space. They are typically part of the Earth's surface, and thus inherit many of its mereological properties. At the same time, however, empirical evidence suggests that geographical objects are organised into categories in much the same way as detached, manipulable non-spatial objects (Mark *et al.*, 1999). Consequently, geospatial data and services contain symbols whose meaning is not only

a matter of convention, but grounded in physical reality. Land use, for instance, has an observable grounding in the world. As the geographer distinguishes between physical and human geography, there are on the one hand physical entities such as mountains, rivers, and other features that make up land cover. On the other hand, there are socio-economic units like cities, neighbourhoods, and land use. Some of those categories are defined by function, for instance a house is a building in which something is sheltered or located (Kuhn, 2007). Geospatial information is often based on human perception and social agreements, combining objective measurements with subjective judgments (e.g. Santos *et al.*, 2007). Providing a mapping between them is probably the biggest challenge to make geospatial information more meaningful and shareable.

Meaning expressed by ontologies provides the long sought for glue between geospatial communities by capturing their practices and conceptualisations, and facilitating the alignment of heterogeneous elements expressed at a high semantic level. Indeed, many high level semantic paradigms have been used to describe the geospatial domain, from image schemata, conceptual spaces, affordances, to multi-representation, and recently difference spaces (Tanasescu, 2007). Logic-based ontologies offer reasoning capabilities about types of geospatial values, objects, and functions. Unfortunately, they do not offer a magic solution to the problem of different unconnected perspectives of different levels of application specificity, or issues relating to handling vagueness as well as cultural and subjective discrepancies (Agarwal, 2005).

According to Freksa and Barkowsky (1995), it is impossible to make all potentially interesting aspects of the world simultaneously explicit within one representation medium. Because the geographic world surrounding us is extremely complex, we usually single out particular aspects of interest from this multifaceted formation. At any given time, we are only interested in few objects and only particular properties and relations. This means to make explicit specific aspects of the world, we ignore others. The ability to switch between views is an important feature of using world knowledge intelligently. For example, Frank (2001 and 2003) suggests that an ontology for GIS should be built as a coordinated set of tiers of ontology, which distinguishes the physical reality and its observations, objects with properties, cultural conventions of the social reality, and subjective knowledge in the form of ideas cognitive agents have about the world. This multi-tier ontology is supposed to recognise that various

approaches contribute to our understanding of certain aspects of the world around us. It is supposed to help with the integration of data from different sources to understand the processes that result in agreement or disagreement between data.

These aspects put high demands on ontology design in the geographical domain. There is the need to share information more easily due to high acquisition and maintenance costs. Spatio-temporal databases must make stronger commitments to capture the meaning of space and time. This requires the modelling of complex spatio-temporal relations such as topology, mereology, and metrics (e.g. Hornsby and Egenhofer, 2000; Raper and Livingstone, 1995). Furthermore, modellers need to account for culture dependent semantics of spatial terms for linguistic as well as professional cultures. However, the complexity of geographically referenced data, the (potentially) large databases to be manipulated and the diversity of application areas make GIS a candidate for the application of artificial intelligence techniques (Miller, 1994). The general objective is to emulate the problem-solving capabilities of the human expert to manage and access data more effectively (Openshaw and Openshaw, 1997). This is the reason why current research focuses on ontologies in terms of interoperability issues, information retrieval, domain specification, knowledge generation and general information system development. As we have seen in earlier chapters, conventional GIS data models suffer shortcomings in the way geographic information is stored and represented, and thus fail to meet specific application contexts. It remains an impossible task to acquire and store all knowledge from raw information before knowledge is accessed, and to provide unprocessed raw information and computing specific knowledge on demand (Freksa and Barkowsky, 1995; Peuquet, 1988; Burrough and Frank, 1995; Frank, 1992). To overcome any of these issues, GIS research must separate the concepts involved in a programme from the mechanics of its implementation as a program (Frank and Mark, 1991). It must separate the conceptual database schema from the physical storage arrangement, while a third schema describes subsets of the conceptual view according to users and their specific task contexts (figure 18).
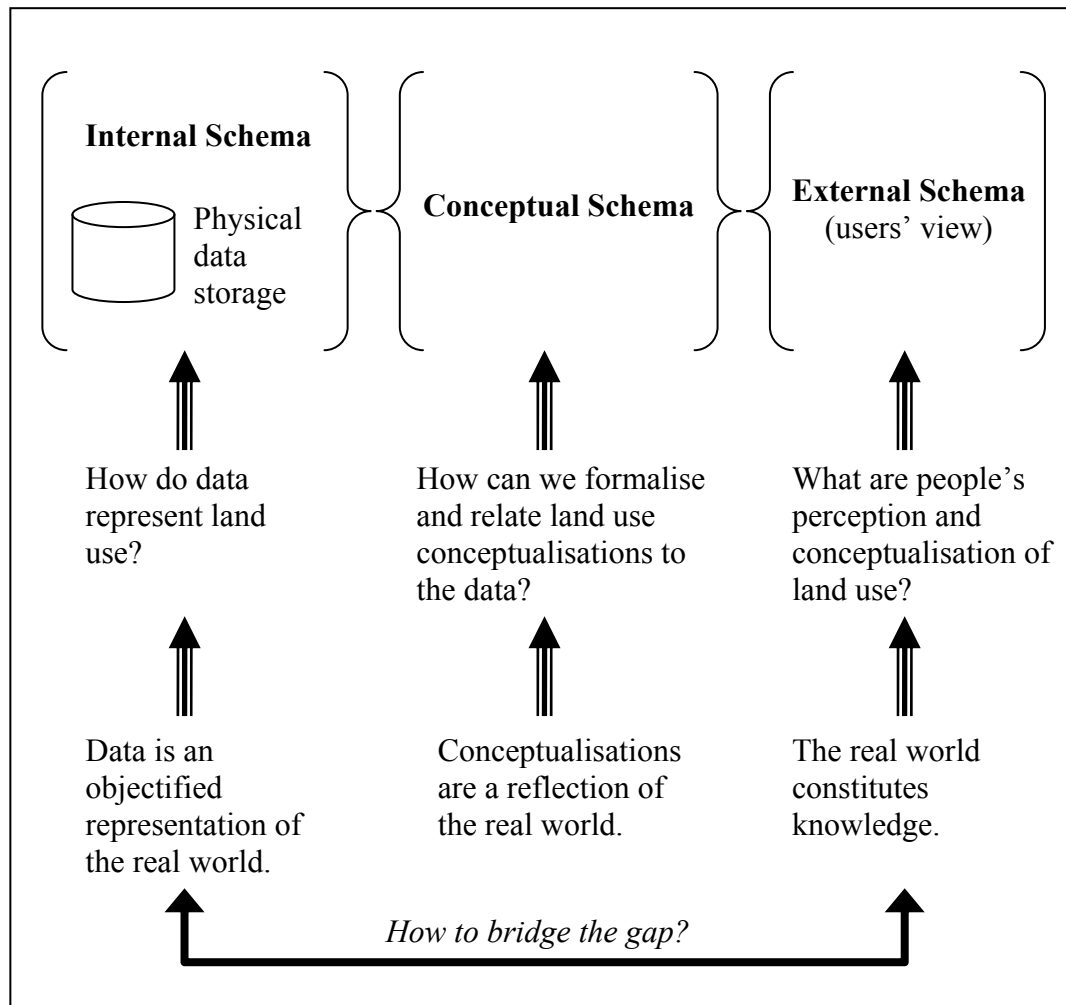
Figure 18 The relation between different schemas

Database schemas constrain focus on data integrity, whereas ontologies constrain focus on intended meaning. We have to bridge the entities represented in a GIS, which stand for the real world objects and their properties, with the ontology, which stands for the conceptualisation of knowledge consisting of some of these real world objects, their relations and properties. Geospatial semantics is not about the relationship between GIS contents and the world. This relationship is already captured in the notion of correctness and integrity of databases and information systems. Geospatial semantics is about understanding GIS contents, and capturing this understanding in formal theories (Kuhn, 2005). A GIS database should therefore present a logical view of the data as well as the derived higher-level knowledge that corresponds to people's own cognitive view (Mennis *et al.*, 2000). An ontology forms the mediating instance between the represented world's reality, i.e., the raw data, and the information that is required according to human understanding about the enquired concepts. In that respect, the ontology becomes a powerful tool for tailoring effective and efficient descriptions of

arbitrary situations from arbitrary viewpoints. As Frank (2003) argues, a meaningful combination of different semantics and representations requires bridging the gap created by ontological assumptions as well as translations between the representations once their meaning is in the same context. However, even for databases where all data are from the same source, the gap between the ontology of the data collectors and the ontological assumptions of the designer of the GIS software and later the users must be bridged.

It is widely accepted that ontologies will play an important role for the next generation of information systems (Wessel and Möller, 2007; Guarino, 1998; Fonseca, 2001; Fonseca and Egenhofer, 1999). Future information systems should be able to handle semantic heterogeneity by making use of the amount of information available with the arrival of the Internet and distributed computing (Fonseca *et al.*, 2003). According to Guarino (1998), the role of an explicit ontology within an information system is to drive all aspects and all components of that information system (IS). This includes both the development and run time of an IS. The use of ontology in an IS component enables the developer to practice a higher level of software reuse than is usually the case in software engineering. The use of a common vocabulary across heterogeneous software platforms helps to concentrate on the structure of the domain and the task, and thus increases the quality of the conceptual analysis process. At run-time, Guarino distinguishes between ontology-aware IS and ontology-driven IS. In the former, the system is merely aware of the existence of an ontology and can use it for whatever specific application purpose is needed. In the latter case, the ontology is another component cooperating at run time towards the overall IS goal. An important benefit for using an ontology at run time is enabling the communication between software agents. An example of an ontology-driven geographic information system is given in Fonseca *et al.* (2000).

As part of the database component of an IS, ontology can be compared with the schema component of a database. Whereas an ontology usually describes a specific domain, a conceptual schema describes the contents of a database (Spyns *et al.*, 2002). The ontology, however, is semantically much richer than a database conceptual schema, and thus closer to the user's cognitive model. This is because conceptual schemas are built to organise what is going to be stored in a database. An ontology, on the other hand, represents concepts in the real world (Fonseca *et al.*, 2003). For example, a common

conceptual schema can be used for mapping heterogeneous conceptual schemas on a common top-level ontology, thus providing access facilities in a heterogeneous environment. Ontologies therefore form the core of the mediation-based approach to information integration. The large-scale towntology project, for example, has been building an ontology to facilitate, in the long term, interoperability between models of databases and different co-operative systems of design, as well as the communication between various actors in the urban management and planning domain (Keita *et al.*, 2004). The fundamental question in interoperability is that of identifying objects in different databases that are semantically related, and then resolving the schematic differences among semantically related objects (Kashyap and Sheth, 1996). Torres *et al.* (2005) argue that GIS-applications require alternative object representations that are independent of the imprecise nature of the data. Montes de Oca *et al.* (2006), for example, use logic rules for the automatic description of spatial data to provide quality information for spatial decision support systems. Ontology offers a useful alternative representation by precisely describing concepts, objects and properties independent of scale, format, or references (Verastegui *et al.*, 2006), thus representing information on a level that corresponds more readily to human cognition. Keßler (2006), for instance, investigates conceptual spaces for data descriptions based on people's perceptions as the fundamental quality dimensions. Gärdenfors' (2000) theory of conceptual spaces provides a mathematical basis for the analogy between concepts and geometric spaces, and has been exploited fruitfully for all sorts of representation and reasoning challenges, in particular for similarity measurements and transformations (Kuhn, 2005). Similarity measurements and semantic matchmaking between concepts is particularly important for achieving interoperability between data that have heterogeneous classification systems (e.g. Rodríguez *et al.*, 1999; Schwering, 2005; Schwering and Raubal, 2005; Feng and Flewelling, 2004; Visser *et al.*, 2000). Ahlqvist (2005; Ahlqvist and Gahegan, 2005), for instance, explores the use of conceptual spaces to translate between taxonomies of land cover categories, and estimates their semantic similarity.

As part of a user interface, an ontology can be useful for mapping a user's natural language terms to the IS vocabulary, thus facilitating improved querying of the stored data (Frank and Mark, 1991). 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). Klien *et al.* (2004) and Schwering *et al.* (2003), for example, extend query capability with terminological reasoning on metadata provided by an ontology-based reasoning component. They illustrate how this approach contributes to solving semantic heterogeneity problems during free-text search in catalogues and on the Web. There has been some work related to optimising queries by connecting ontologies to existing relational databases (e.g. Beneventano *et al.*, 2003; Calvanese *et al.*, 2006a; El-Ghalayini *et al.*, 2005). More recently, Zhao *et al.* (2008) try to overcome limitations of using ontology-based queries that yet cannot be applied directly to legacy data stored in databases by rewriting user queries as SPARQL queries. Ordnance Survey also invests research in this area to map spatial domain knowledge directly to databases (Dolbear *et al.*, 2007). To overcome the mathematical nature of description logics that makes it difficult for non-logicians to understand and author logic-based ontologies such as OWL, Ordnance Survey developed a Controlled Natural Language syntax called Rabbit, which is an engineered subset of a natural language with explicit constraints on grammar, lexicon, and style (Schwitter *et al.*, 2008). Other work tries to improve the user interface by detecting mismatches between a user's and an expert's conceptual model (Huang *et al.*, 2005). Peachavanish and Karimi (2007), for instance, try to mitigate the knowledge gap between non-experts and experts in GIS by using ontological-based methodologies and techniques to automate tasks related to the interpretation of geospatial queries and mapping the interpreted results into geospatial data models and geo-processing operations. The use of conceptualisation of geospatial queries, knowledge representation for queries, and ontologies to map queries to geo-processing operations shall aid non-expert users to solve geospatial problems with little knowledge and skill on the workings of GIS platforms. However, a successful method for expressing queries is still needed.

As part of the application program component, it is possible to represent explicitly all the domain knowledge that is hidden in the application program, thus turning the program into a knowledge-based system. Verastegui *et al.* (2006) for example, incorporate semantic content into a spatial database to support subsequent processing. Their method is based on conceptually representing topological and geometrical properties of the data, which are only implicitly contained within the database. The creation of a knowledge-based system allows the inference of new knowledge, adding it to the database (Montes de Oca *et al.*, 2006). As we learnt in chapter 2, the

augmentation of cartographic and other geodata with semantics supports tasks such as generalisation and improves usability. In fact, semantic enrichment has been long discovered as a promising application area for well known AI-techniques and methods. Its adaption to the geospatial domain means that we need to attach semantics to information sources and entities, and to draw conclusions from the semantic annotations available (Visser *et al.*, 2000). This kind of intelligent information processing is what makes ontology so relevant to this thesis. It provides a means to discover and retrieve implicit information from geospatial databases. Geographic data models tend to explicitly represented only a set of basic objects, their geometry and their properties (Worboys, 1996). However, much of the semantics appears in the relations linking objects. Some relations are represented in data models, others are not. The extraction of implicit information is based on the unique set of characteristics that are inherent in the data. These characteristics encompass geometric and other unary properties, spatial relations such as topology, distance and direction, and existing attributes, all of which are easily recognisable by humans, but are mostly contained implicitly within the dataset. Often they need to be calculated and derived through a number of GIS operations. However, even upon finding implicit spatial structures, the computer still does not know the meaning of them – the semantics needs to be translated into a computer-comprehensible way (Heinzle and Sester, 2004).

There is enough evidence to pursue the use of ontologies to infer functional information from topographic knowledge with an optimistic view on success. In particular, Barr *et al.* (2004; Barnsley and Barr, 1997) investigated the intrinsic separability of several different categories of urban land use based on the morphological properties of, and the spatial relations between, their component land-cover parcels (see chapter 2). They performed a statistical separability analysis to validate their assumptions and to provide quantitative evidence. Unfortunately, Barr and Barnsley have not taken their developed extended relational attribute graph (XRAG) to the next level of actually searching for land uses that meet their a priori established morphological characteristics. Klien and Lutz (2005), on the other hand, illustrate an automated semantic annotation of geodata by associating spatial analysis methods with spatial relations in ontologies. By establishing concept definitions for the intended domain ontology, and extracting each concept's characteristic spatial relations, it then becomes possible to analyse an existing non-annotated dataset for spatial entities that meet the specified characteristic spatial

relations of that concept, thus annotating it as such. On the one hand, GIScience's multi-disciplinary nature lies at the root of many of the semantic problems. On the other hand, it offers methods in ontology design that are informed from development across various disciplines (Agarwal, 2005).

The transition from data, to information, and to knowledge is achieved by defining a conceptual model that offers expressive facilities for modelling directly and naturally, and for structuring information bases. In the simplest case, an ontology describes a hierarchy of concepts related by subsumption relationships. In more sophisticated cases, suitable axioms are added to express other relationships between concepts and to constrain their intended interpretation (Guarino, 1998). Ontologies offer benefits in terms of ease-of-maintenance, extensibility and flexibility, and they can help to increase the transparency of application software. The semantic database industry, with examples such as Freebase and True Knowledge, has already developed technology that demonstrates the benefits of semantic databases (Lowe, 2008). Benefits such as having a much richer, structured data modelling approach have been known for a long time (Peckham and Maryanski, 1988; Hull and King, 1987). However, who defines and categorises data into these types and who builds the relationships between database elements? Knowledge acquisition can indeed form a bottleneck in this matter (Sester, 2000). Even with the wiki approach that Freebase uses, the question remains whether it will scale up. Semantic databases will become the future in the way we interact with information only when their development and maintenance can become automated. Ontology, however, is merely a theory of objects and their relations, concerning especially entities in language. It offers an explicit specification of a conceptualisation, which can be formalised in a machine-readable way. A common misconception is that ontologies are a collection of facts arising from a specific situation. An ontology is more than just facts, it defines intended meaning of these facts. By itself, ontology is not a database schema or a model of an application domain, nor a vocabulary or dictionary, not even a knowledge base. Ontology is the general framework for organising knowledge. Its defined concepts therefore can become part of a domain model or a knowledge base, and this thesis explores this in the subsequent chapters. Ontology may offer human-legible and structured content that assists with interoperability and mixing of data and metadata. However, ontology languages are not designed to cope with context other than by building other ontologies and producing mappings between them

based on syllogistic reasoning (Tanasescu, 2007). They require categorisation that is often arduous and fails to reflect the dynamic nature of the geographic environment. The derived symbol-based systems often are not grounded in reality. They are separated from subsequent design products or are not accessible to application programs, as in the case of querying databases. Furthermore, ontology can have very long descriptions. It is extremely difficult to encode very large geographical databases. There is no indexing, which makes querying within a knowledge base very memory-dependent and slow. As a result, some regard ontology as a panacea for an extremely wide range of problems. Others adopt entirely the opposite view and deny that ontological inquiry has any sense at all. On the one hand, there is the desire to capture the world in a definitive set of categories. On the other hand, there is the opinion that the endeavour is impossible and nonsensical (Poli, 1996).

In spite of both baseless enthusiasm and deligitimating rejection of ontology and all of its controversies, research is advancing in this area and ontologies are becoming a widespread phenomenon with the semantic web. In the light of this thesis, I can discern three major advantages of pursuing an ontological approach for three respective viewpoints: From a business perspective, the semantic enrichment of OS MasterMap, for example, will be useful for its wider use, meeting a wider range of customer needs, and offering customised, thematic representations. Ordnance Survey is particularly interested in describing their data to the user's understanding, as well as to facilitate data integration and ontology merging. From a customer's perspective therefore, this means that data are potentially more suitable for the way customers solve their problems. This includes new ways of cartographically drawing the information to customised maps. From an academic perspective, the thesis distinguishes itself from other research by exploiting ontologies for inferring higher-level knowledge from topographic data, and by tying knowledge representation closely to the way people interpret topographic maps.

4.2 The ontology engineering approach

How is ontology useful for this particular problem of exposing high-level semantic information within topographic data? According to Nunes (1991), geographic space is a definition of its geographical objects, their attributes and relationships. Ontology is ideal

for modelling these entities as well as human-legible, high-level concepts that describe land uses and functional information. Because this knowledge is not explicitly contained within the data, the goal of this thesis is similar to the broad field of data mining and knowledge discovery. Knowledge discovery in databases (KDD) is a process of identifying valid, novel, useful, and understandable patterns in data (Fayyad *et al.*, 1996). Most literature however is about validity and process and very little is about novelty, utility, and understandability. Psychological studies of the nature of comprehensibility of knowledge structures are necessary to give substance to the intuitions. For Pazzani (2000) it is time for KDD to draw on cognitive psychology in addition to databases, statistics, and artificial intelligence. Pazzani argues that by considering the human cognitive processes, we can increase the usefulness of KDD systems. This does not suggest that KDD systems should emulate the way people learn from data, since people have difficulty finding subtle patterns in terabytes of data. However, KDD can benefit from incorporating some of the human learning biases.

Ontologies play already a role in KDD systems, but mostly as background knowledge. They express the main concepts and relationships in a domain in a way that is consensual and comprehensible to the given professional community, committing to some generic principles of knowledge organisation. Their role depends on the given mining task and method, on the stage of the KDD process, and on some characteristics of the domain and dataset. Usual applications of ontologies are data understanding, task design, result interpretation and result dissemination over the semantic web (Svátek *et al.*, 2006). Although ontologies are a popular instrument in many diverse applications such as text mining (e.g. Vallet *et al.*, 2005), they have been mainly used to enhance the knowledge discovery process. This thesis, however, will explore the reasoning abilities of logic-based ontology languages for the actual mining process.

Discovered knowledge should be concise, informative, and be represented by high-level concepts. Ontologies present themselves as ideal candidates for modelling these semantic requirements considering their main characteristics (Kuhn, 2004a):

- Semantics has a model-theoretic view of the world, assuming that the meaning of an expression is a model of the real world.

- Semantic theories are axiomatisations for predicates in some variant of first-order logic, for instance in a description logic. Hence, they are machine-interpretable.

- All formalisms assume compositionality, that is, the idea that the meaning of a component is the set intersection of the meanings of its components. This is particularly useful for describing land use information that is naturally compositional.

- The meaning of sub-types is explained by adding predicates to account for additional information, such as the sub-type house to an ontology describing types of dwellings.

- Multiple inheritance relationships, which are essential to any theory of meaning, can only be handled if the sub-type stands for the set intersection of what the super-types stand for. Therefore, instances of the sub-types inherit all attribution from the supertype.

These characteristics will become clearer in due course as we develop a semantic model for exposing functional information within topographic data. The question is how logical descriptions assist in the tasks of geospatial data interpretation, or map description (e.g. Montes de Oca *et al.*, 2006), and especially in interpreting functional information from topographic maps. The key is the reasoning facilities provided with logic-based ontology languages, which chapter 6 discusses in detail. Reasoning in that sense relates to the psychological approach of deduction, where reasoning is a kind of mental process that creates new ideas from old ones (Rips, 1994). There is not only the need to infer new information from existing ones, but to classify database instances according to some defined high-level concepts. Classification is a well-known data mining technique, where the data stored in a database are analysed to find rules that describe the partition of the database into a given set of classes. Each object in a database is assumed to belong to a predefined class. The most common classification method constructs decision trees that use a top-down, divide-and-conquer strategy for partitioning a set of given objects into smaller subsets (e.g. Koperski *et al.*, 1998). With ontologies, we can devise a similar hierarchical tree of defined super- and sub-types of concepts, and classify instances accordingly.

The first step towards bridging the divide between existing models of space and the way humans interact with and conceptualise space is to make higher-level semantic information accessible from existing data repositories. The crucial point is therefore the incorporation of the human component. People effortlessly combine contextual and configuration information with other significant variables such as size, shape, similarities and proximities to draw inferences, as chapter 3 elaborated. Our inferences happen so quickly that we are often not aware about the 'why' and 'how' we came to our conclusions. By making this knowledge explicit through ontology – hence modelling the relevant spatial relations, typical characteristics and other variables common in the geographical space – we conceptualise and translate this knowledge into machine readable and logical form. The resulting formal ontology then becomes a useful tool for:

1. Modelling all the aspects that are relevant for a complete representation of some functional information concepts,

2. Deriving previously unknown (functional) information from the database as part of the data mining process,

3. And simplifying the process of sharing and integrating the existing database content with other data sources.

Consequently, ontologies have the potential to bridge the gap between a given conceptualisation and the data it relates to (e.g. Hart, 2007). They can be seen as a mediating instance between the captured reality, i.e., spatial data, and higher-level knowledge.

Unfortunately, a standard, unified and acknowledged methodology for the building of ontologies is still missing. Figure 19 is a combination of several proposed approaches in literature, whose methodologies have general stages in common (e.g. Jones *et al.*, 1998; López, 1999; Noy and McGuinness, 2001; Mizen *et al.*, 2005; Kovacs *et al.*, 2006). A form of ontology life cycle appears with support and development-oriented activities. The starting point is an initial specification of the ontology, including its motivation, purpose, scope and domain. This is followed by knowledge acquisition to capture the main concepts, properties and relations of the ontology. The questionnaire survey in chapter 3 provides such knowledge for the ontology development. The next stage, conceptualisation, addresses the informal description of the ontology. These are usually natural language descriptions, or expressed as graphical diagrams like semantic

networks (e.g. Kovacs *et al.,* 2006). The next chapter defines the conceptual framework and requirements for this thesis. The informal description then carries through to the formal description, where concepts and relations are expressed as axioms, and definitions are stated in a logic-based ontology language. Chapter 6 introduces the necessary formalisms, which are essential for the implementation of the ontology in a knowledge-based system. The ontology is then ready for application to whatever particular problem it addresses. The thesis applies the developed ontology to topographic data for deriving implicitly stored functional information (chapter 7). During ontology evaluation, the ontology is assessed in regards to its structure, intended use, and usability. Chapter 8 draws conclusions about the usefulness of ontology in terms of its design and reasoning abilities. Either the circle then closes and the ontology becomes a matter of maintenance and possible future re-use, or the circle rejoins the specification stage because there are re-adjustments in the ontology's conceptualisation as well as formalisation.
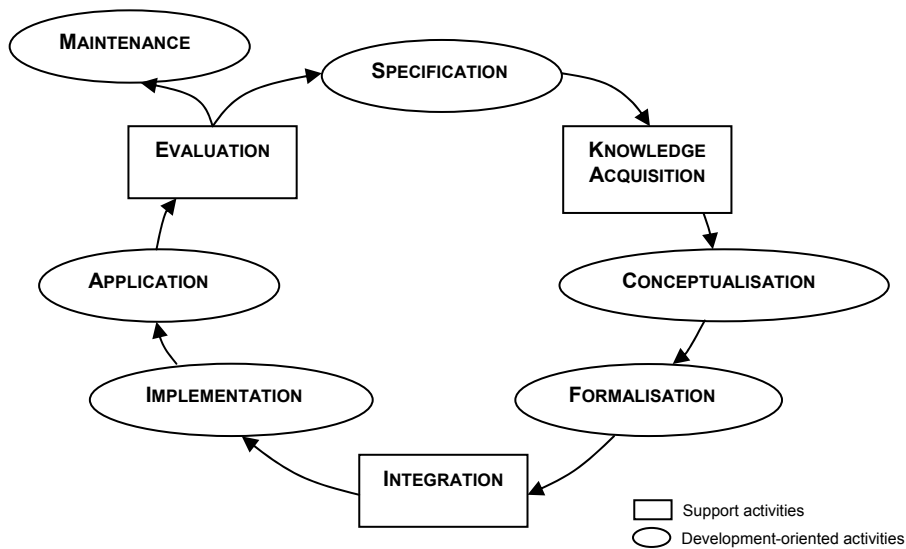
Figure 19 Ontology development and life cycle

Just as ontologies are classified according to their level of formality, i.e., informal, conceptual, and formal ontologies, they are also distinguished based on their level of granularity. Granularity is expressed through vertical neighbours of sub-concepts that describe a certain level of detail. For the ontology to become dynamically adaptable to different situations, vertical neighbours must either unify or split as a means to decrease or increase the level of granularity respectively. For example, a coarse ontology consists of a minimal number of axioms and is intended to be shared by users that already agree

on a conceptualisation of the world. A fine-grained ontology needs a very expressive language and has a large number of axioms. Often the terms high-level and low-level ontologies are used respectively (Fonseca *et al.*, 2002). This distinction stems from the ontology integration problem (Guarino, 1998). The need to find an agreement between different models of conceptualisation has led to the consideration of developing different kinds of ontologies according to their level of generality. For our purpose, granularity must meet the level of detail provided by the objects, attributes, and relations of geographical space, that is, of the data under interrogation. If we want to map individual features of the data onto the higher-level conceptualisation of the ontology, we require a sufficient vertical level of granularity. This is what differentiates this approach to many existing ontologies that were designed primarily for data sharing. For example, the land use ontology of the HarmonISA project provides just enough formalisation to describe relevant categories according to some general attributes (Hall, 2006; Mandl and Hall, 2006).

A classification according to an ontology's dependence on a specific task or viewpoint has been proposed to differentiate between the different levels of granularity (Guarino, 1998; Fonseca *et al.*, 2002). Figure 20 illustrates how concepts are interrelated between these different levels of generality. They range from the fine detailed, low-level data ontology (describing the data), via the high-level concepts of the domain of interest (describing the functional information) to an existing upper level ontology that provides the most generic concepts and relations for classification. The top-level or upper ontology is an attempt to create a unified ontology, which describes general concepts that are the same across all domains, thus providing concepts and relations to the more specific domain ontology (Guarino, 1998; Masolo *et al.*, 2003; Niles and Pease, 2001). At the domain level, the vocabulary relates to a generic domain, as in this case of our high-level functional information. Sometimes the ontology may refer to a specific task or activity, such as inference or information retrieval, for which the terms introduced in the level above are further specialised. Application ontologies are a specialisation of both the domain and task ontology, where concepts often correspond to roles played by domain entities while performing a certain activity. By putting domain ontologies on the foundation of an upper level ontology like DOLCE, this potentially helps to enhance the quality of the domain and application as well as to achieve not only logical consistency but also ontological consistency (Klien and Probst, 2005). Because of the limited scope

of this thesis, I will concentrate on the application level. As long as the detailed ontologies are based on high-level ontologies, so that each new ontology level incorporates the knowledge present in the immediate higher level, a mapping can be achieved between these different levels of abstraction (Fonseca *et al.*, 2002).

```
Top-Level      ┌──────────────────────────────────┐  [Philosophers &
               │ Upper Level Ontology (e.g. DOLCE, SUMO) │  cognitive engineers]
               └──────────────────────────────────┘
                        Provides concepts and relations to

Domain-Level   ┌──────────────────────────────────┐  [Domain experts]
               │ Domain Ontology: A conceptual framework for │
               │ defining higher-level functional information │
               └──────────────────────────────────┘
                        Provides concepts and relations to

Task-Level     ┌──────────────────────────────────┐  [Service]
               │ Task Ontology: Inference of functional │
               │            information            │
               └──────────────────────────────────┘
                        Provides concepts and relations to

Application-Level ┌──────────────────────────────────┐  [Specific Example]
               │ Application Ontology: model & implement │
               │ the high-level concept residential │
               └──────────────────────────────────┘
                        Link

Data-Level          OSMM  - - - -> Map1    [On-demand mapping]
                                   Ultimately leads to
                                   new representations
```

Figure 20 Ontologies at different levels of generality

Ontology engineering can take either the form of a top-down or bottom-up approach. With the top-down approach, one defines high-level concepts first and then adds specific concepts from the ones that are more generic. The potential difficulty here lies with the inclusion of low-level, real world objects. The bottom-up approach, on the other hand, starts from real existing objects. It aggregates them into more generic objects, and finally arrives at the top-level concepts. A compromise is the middle-out approach, where concepts are both generalised and specialised (Freksa and Barkowsky, 1995). Overall, the ontology forms an organised structure within which catalogues, taxonomies, terminologies may be given suitable organisation. The resulting structure is a network that represents knowledge through links between the individual concepts constituting and stating their relationship to one another, whether it is a part-of relation, topological relation, or any other. By organising the knowledge in this way, natural associations of the concepts can be included in the ontology. The more inclusive and

abstract categories, for example the core concepts of functional information, reside at the top, with narrower and more specific categories, mostly geographical objects, beneath them. As in any network, it also becomes possible to assess neighbourhood relations among concepts, that is, all core-, secondary- and sub-concepts. The horizontal neighbourhood between concepts refers to competing concepts on the same level of granularity, allowing for the selection of an appropriate description value. Concepts that are far apart in terms of the distance along the network will have less in common than ones close by. At the vertical level, neighbouring concepts reveal their part-of and kind-of relations, and refer to compatible concepts on different levels of granularity allowing for the selection of an appropriate description granularity. The resulting knowledge representation, or concept space (Freksa and Barkowsky, 1995), is therefore not simply a hierarchical tree subdivided by individual concepts, but rather constitutes a neighbourhood structure for objects, relations and attributes of various kinds. This property allows the capture of implicit knowledge about conceptual aspects from the description of other concepts, because a concept's meaning is largely given by their relation to other concepts.

A mutual relatedness between concepts and objects enforces a certain structure upon the concepts. As established earlier, space intrinsically ties geographic objects to its structural properties leading to special ontological considerations. Concepts have their meaning rooted in their relation to objects. For example, the concept of a building rests on the object's properties in regards to its size and shape. Considering concepts in their wider context requires spatial relations. For example, a terraced house is defined by its relation to its neighbouring buildings. A spatial conceptualisation therefore must be able to contain a qualitative topology, that is, a theory of boundaries and interiors and connectedness and separation, as well as a mereological theory of parts and wholes, and qualitative geometry (Smith and Mark, 1998). Modelling these relations is potentially very difficult. Whereas we are intuitively aware of an objective reality that contains so-called bona fide objects, such as buildings, lakes, and roads, the human geographic reality includes also objects that exist only in virtue of our individual and social conceptualisations of the relevant areas of space (Peuquet et al., 1998). Classical problems connected with the notions of adjacency, contact, separation and division may be resolved in an intuitive way by recognising a two-sorted ontology of bona fide and fiat boundaries. However, Smith and Varzi (2000) argue that their opposition cannot be

modelled in a natural and intuitive fashion within a topology on a set-theoretic basis. Therefore, if we force a conceptualisation into logical formalisms, we have to accept that the resulting categorical system exists only because of delineating boundaries between categories, even if reality is imprecise, such as land use, where boundaries are partly induced through human demarcation.

To the contrary, the meaning of spatial concepts is not only given by their relations to physical objects, but by their relations to other concepts. For example, we can imagine the concept residential through other concepts such terraced houses or semi-detached houses. The compositional aspect of these kinds of concepts shows that meaning is constituted even further through situation context. In this sense, Freksa and Barkowsky (1995) argue that spatial concepts have a meaning independent of an envisioned physical materialisation. This independence is what makes concepts universally valid for all situations in which the concept can be used. However, I believe that a concept's meaning is always stimulated by the comprehended reality of the environment we live in. Consider the land use domain that contains such abstract, universal concepts like residential or recreational. Does not a functional concept become spatial through its conceptualisation in terms of its geographic space it claims in the real world? Such a concept is made up of real geographical objects, such as a building where recreation can be practised. This means, a functional concept comes into existence based on its part-of relations that constitute its spatial context and configuration. Consequently, a concept is defined through its envisaged physical and spatial existence whether this is expressed through other concepts or objects.

Consequently, we need to match the discrete world of concepts with the continuously perceived world of features of the entities in the real world. This includes both categorical predications, such as building, house, or road, and accidental predications (properties), such as large, living, or natural (Mark *et al.*, 1999). We need to capture geographic objects according to their surroundings and context, such that for instance a land parcel of sand only becomes a beach if it is adjacent to the sea. Geographical objects may be persistent in space and only change very slowly, however their associated functions potentially change much more rapidly. A building may be used as a residential home one day; another day, it may be transformed into a set of offices. I will only consider the static viewpoint, the situation given at one time. In chapter 3, part of

the questionnaire survey generated instances of categories and attributes at several levels, evaluating their goodness and typicality. There is a great degree of agreement among human subjects as to what constitute good and bad examples of category members. Naturally, we learn how the things falling under given categories are related to each other and how they interact causally, but when we want to model a category system, we must make concrete decisions on their relationships and definitions. The elicitation of ontologies not only helps to make these decisions, but also to model domains according to the conceptualisation of given individuals or cultures. Instead of focusing on knowledge and beliefs in general, an elicitation concentrates on the ontological content of certain domain-specific representations. Considering the need to model both low-level, object-specific content as well as high-level content, a middle-out approach that combines both top-down and bottom-up approaches seems most suitable. Thereby we will go top-down from the concepts established from the human knowledge and natural language descriptions towards geographical objects, and bottom-up from the objects towards the concepts. Although both methods will reach the same level of granularity at some point, Freksa and Barkowsky (1995) warn for incompatibilities when pursuing a one-to-one mapping with this approach.

This duality is caused by the nature of our problem. On the one hand, we are dealing with human understanding, meaning and knowledge of land use, whether derived from questionnaire surveys or existing natural language definitions (e.g. Fellbaum, 1998). On the other hand, in a GIS we are dealing with scientific fiat boundaries, prescribed by the classification system of the spatial representation. The entities stored in a database are mathematical fiats that are artefacts of a certain technology. These artefacts can have measurable quantities and other physical properties that constitute fields that vary more or less continuously and somewhat independently across geographic space (Smith and Mark, 1998). With this kind of perfect information, we can apply a typology of land use to the continuous variables and derive crisp boundaries. Hence, land use can be seen as a world of geographic objects with crisp boundaries. They may misrepresent the phenomenon, but they are the best that can be done with current data representations. Consequently, the ontologies underlying most geographic information systems rest on discretised metric world models. On the contrary, a high-level ontology must have the resources to represent the qualitative conceptual categories conveyed by natural language. Therefore, a mapping of high-level concepts onto the low-level features is

required. An ontology allows us to combine both types of knowledge granularity. Further, by formalising the ontology with an appropriate knowledge representation language, knowledge becomes machine interpretable and can be used to infer new knowledge. This process is similar to the goal of data mining, as discussed earlier. Figure 21 depicts the KDD's general basic steps from understanding the application domain, its relevant prior knowledge, identifying the overall goal, to searching and interpreting mined patterns, and consolidating the discovered knowledge (Fayyad *et al.*, 1996). These general processes are pliable to our specific problem. Therefore, the figure can be altered to show exactly where ontology slots into the KDD process and how the information retrieval procedure will be performed.
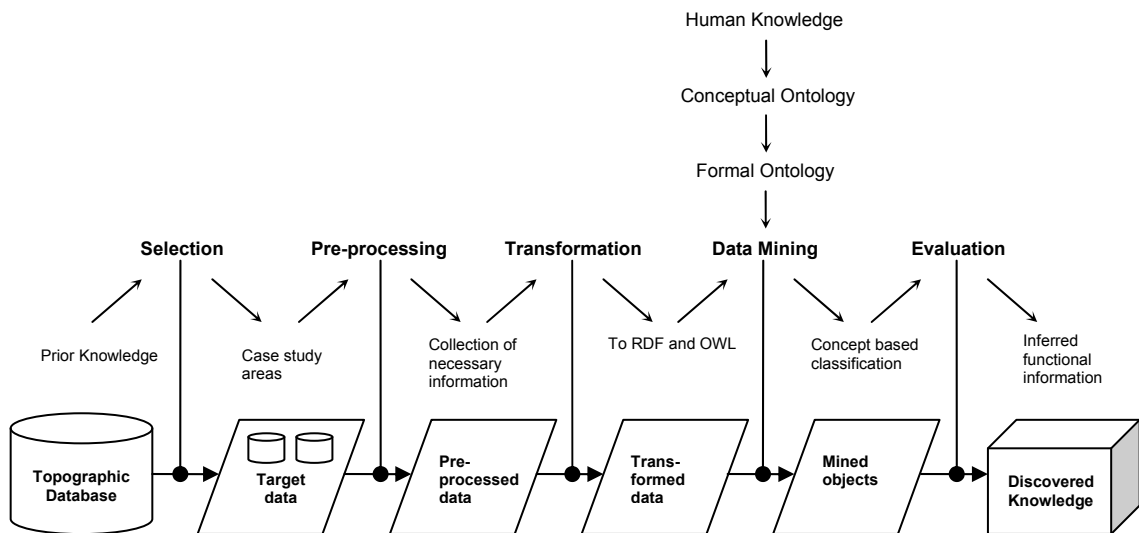
Figure 21 Ontologies for knowledge discovery

The starting point is the topographic database – in data mining lingua often referred to as prior knowledge. From this large repository of data, we select some sample data. This data then is pre-processed to collect all the necessary information, such as spatial relations, building features, and other useful attribution. The transformation stage relates to the reduction of the target data to the ontology language RDF/OWL, so that the information can be imported into the knowledge base. The knowledge base consists of both the asserted knowledge about the data as well as the high-level concepts described by the ontology. The ontology serves as a tool for information retrieval by classifying the asserted topographic instances according to its high-level concepts. After the concept-based instance retrieval, the inferred functional information is finally added as new knowledge to the database. The next chapters will further explain and illustrate this

crudely outlined method with examples from OS MasterMap Topography Layer. This solution is based on the cognitive use of a priori knowledge to interpret and categorise new observational data. It rests on the compromise of pre-processing raw data to a level of conceptually relevant granularity and using this pre-processed data on demand by the conceptualisation formed in specific contexts (Freksa and Barkowsky, 1995). Consequently, the suggested approach exhibits a strong correspondence to the human use of geographic maps.

Conclusions

Ontology is one of the most thriving themes in geographic information science. In the past, ontology has been rather confined to the philosophical sphere, but now it has become a fancy name for its role in AI, computational linguistics and database theory (Guarino, 1998). As a result, there have been many ambiguities around the term ontology (Guarino and Giaretta, 1995). It is referred to as a philosophical discipline, as a conceptual system, a specification of a conceptualisation, a representation of a conceptual system via logical theory, a hierarchical structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base, and so forth. Nevertheless, what all these views have in common is that ontology aims to make sense of what exists.

From a philosophical point of view, ontology accounts for a certain vision of the world and makes assumptions about the meaning of terms describing this world explicitly. In AI, ontology has evolved to an engineering artefact that models and represents knowledge of the real world by using a systematic and axiomatic development of the logic of all forms and modes of being. However, ontologies have their own methodological and architectural peculiarities. From a methodological perspective, it is the highly interdisciplinary approach to analysing the structure of a given reality at a high level of generality and in formulating a clear and rigorous vocabulary. The problem here lies with capturing a conceptualisation that is possibly infiltrated with biased knowledge and erroneous beliefs, especially if taken from human subjects. On the architectural side, it is the centrality of the role that an ontology can play in an information system, leading to the perspective of ontology-driven information systems. Here, we need to be vigilant to model a conceptualisation that is broad enough to

capture high-level concepts, but also fine-grained to account for the level of detail given in low-level representations. Therefore, research on this topic must be careful to distinguish the domain of the real world from the domain of computational and mathematical representations, and both of these from the cognitive domain of reasoning, language, and human action (Smith and Mark, 1998).

As it is debatable whether or not the current object-field representation dichotomy of spatial data can represent all kinds of geographic phenomena, ontologies have to encapsulate not only the meanings linked to specific concepts in the data but also the way these meanings are handled and represented in the cognitive set of individuals (Agarwal, 2005). Ontology provides an opportunity to understand better and in a systematic manner geographic reality and the way such knowledge is, or better should be represented in modern information systems. Ontology by definition, attempts to clarify and set the explicit knowledge of the domain they describe. There is a higher ontological perspective with an interest in representing appropriately reality (geographic in our case), or more precisely our knowledge about reality. On the other hand, there is the lower design and implementation perspective with an interest in formalising, processing and associating existing information or data (Kavouras, 2003). For the objective of this thesis, I am proposing a middle-out approach for the ontology development. It seems most promising to combine both interests, top-down from the human elicited conceptualisation of the land use domain and bottom-up from the representation, or more precisely the objectification, of topographic data. Not only do we need to model spatial context of features within the topographic map, but this knowledge needs to be put into a machine-readable format with which we can reason about the data and derive new knowledge in an automated way.

It is evident that the specific nature of geographic categories and the predominantly cognitive nature of geographic information make it difficult to organise the domain and the concepts in it within a structured formal framework (Agarwal, 2005). Even if the organisational structure of such a framework resembles Rosch's (1978) proposed mental category system (chapter 3), capturing the meaning of a conceptualisation in a rigorous way is not easy. Individual conceptualisations are highly subjective and dependent on context. The challenges that we face in the light of the geographic domain are in fact well known (Freksa and Barkowsky, 1995): We have to handle knowledge

that is to be used for different tasks requiring different resolutions and different conceptualisations. We have to manage incomplete and imprecise information. Knowledge and concepts of varying granularity cause fuzzy correspondence between concepts and real world entities. We have to deal with complex open worlds whose dimensions and values cannot be entirely specified. Furthermore, the number of ontologies developed is not large and their practical use in final and real applications is still small. This is especially true for the geospatial domain (Klien and Probst, 2005). However, Kuhn (2004a) believes that these problems generally are tractable. For example, feature attribute catalogues, conceptual data models, descriptions of work procedures, and other sources are subject to explicit and documented agreements among their designers and users. These agreements form the basis of information system semantics, and are, though often implicit and imprecise, available for inspection. They can be mined in the process of defining ontologies. What remains, therefore, is the integration of the information system's semantics with the cognitive nature of geographic information in a framework that provides access to the rich, higher-level knowledge of people.

Chapter 5

A Conceptual Framework for Accessing Knowledge

"Geographic space must be a relative space, and because the concept of relative space essentially means that objects are the space, the problem of defining geographic space or a conceptual model for it is actually a problem of defining and studying the geographical objects, their attributes and relationships."

–Joan Nunes (1991, p.27)

Exploring, learning and understanding visual scenes are a matter of breaking down an image into the sensory input of features and objects that compose a scene description. The key is to understand the concept of what is being recognised, that is, the meaning of things rather than just their appearance. Understanding the meaning of topographic features is also more important than their mere appearance in a geo-referenced space. According to Schröder (1999), if the aim of a scene interpretation is to be in a human understandable form, then the interpretation must operate within terms that the human mind created for the phenomena in the world he or she lives in. Now, how can we recognise meaningful concepts from topographic data in an autonomous way? Let us recapitulate what we found out so far. Firstly, the interpretation of land use information can be defined as a configuration problem because the functional meaning is inherent in the spatial constellation of its land cover features. Secondly, the interpretation is based on map cues such as feature sizes, shapes, and proximities, which can be described with spatial relations. Thirdly, an ontology provides us with a framework for modelling this knowledge in an explicit way. Its different levels of granularity allow us to map low-level, detailed information onto high-level, aggregated, more meaningful information. We have seen that reasoning refers to the cognitive, computational and formal aspects of making logical inferences about a spatial environment (Worboys, 1995). Any form of knowledge retrieval therefore requires powerful inference capabilities.

Through model theory, formal semantics introduces the notion of possible models that are considered to be the meanings of things. Although formal semantics represent meaning as a relationship between symbols of a language or symbols representing

concepts, these relations do not exist in the real world. Rather, they exist in minds to aid in making sense of the world and in interacting with it (Mark and Frank, 1996). From a conceptual point of view, semantics are just symbolic structures, but useful ones: They represent conceptualisations of entities, properties, and relationships in a domain and can therefore be tested against human intuitions. The closer the models correspond to the human concepts about a domain, the more useful will an ontology be (Kuhn, 2005; Jakulin and Mladenić, 2005). A knowledge base incorporates these conceptualisations by creating a unified, high-level collection model (Lewis and Sparck Jones, 1996). It provides more depth and integration through an organised superstructure over the data, hence allowing more intensive inference. Indeed, there are existing methods for describing spatial data, and using knowledge bases and logical rules to infer new knowledge. However, this only happens at the data level. These approaches do not consider how to derive higher-level classes of information (Montes de Oca 2006; Mullally and O'Donoghue, 2006). The thesis is interested in the representation of context in a model that potentially helps to solve semantic problems of similarity and inference at a high level. There are computational benefits that might accrue in modelling and representing context in AI and knowledge-based systems (Kashyap and Sheth, 1996). In a manner akin to database views, a conceptual model provides a semantic summary of the information stored in the data (explicit or implicit). It provides flexible semantics in the sense that the same two objects can be related to each other differently in two different contexts. Furthermore, logical reasoning in a knowledge base easily identifies any inconsistencies inherent in the semantic descriptions.

This chapter introduces the generic framework of how new knowledge can be made explicit within topographic data using a conceptual model. The model expresses the higher-level semantics we wish to recognise in the data. Its hierarchy of concepts captures the semantic distinctions necessary for generating land use information in topographic data. These distinctions require different levels of semantics to link low-level representations, such as house, with high-level representations, such as a district of houses or residential area. The semantics are therefore abstract beyond the data, but broken down into their constituting lower-level features, they still maintain a commitment to the data as a motivating force. In the next sections, I will explain how this conceptual model is built and how it ties concepts from the questionnaire survey to the data through semantic aggregation. We will learn how computer vision motivates

the interpretation of scenes, and more importantly of spatial data. The concept residential area serves as the illustrating example. It is the most straightforward concept people recognise within topographic data. In addition, the spatial patterning of housing yields important information (e.g. Pesaresi and Bianchin, 2000), thus making 'residential' a relevant and important concept in real world applications.

5.1 Decomposing the link between knowledge and spatial data

The entities represented in a GIS stand for the real world objects and their properties. In that sense, GIS forms the mediating instance between world's reality and the way humans interact with this reality. However, when people describe certain aspects about the real world, we have to match the discrete world of concepts with the continuously perceived world of features of the entities in the real world. By using concepts we can see the world as detailed as we need, or in other words, by using concepts we construct our reality of interest (Freksa and Barkowsky, 1995). Naturally, spatial concepts are essential for representing knowledge about the geographic world. A geographic feature has therefore a two-fold meaning: It is both a real world geographical entity and it is a digital representation (Tang *et al.*, 1996). In its digital representation, a feature is committed to the conceptual schema of the computer representation. The data primitives in a GIS contain attributes and relations about the spatial and non-spatial components of the feature. However, a feature is also an instance of an entity set, where the entity is a real world phenomenon. In this sense, a feature is committed to a knowledge domain of an ontology that describes concepts in the way people perceive them (Fonseca *et al.*, 2003). The aim is to allow the user to access information stored in databases using high-level concepts – in this case relating to land use information. This means we have to make the semantics embedded in the geographic data primitives explicit and relate these to higher-level semantics as described by the knowledge domain.

In computer vision, the interpretation of a visual scene is viewed as an information-processing task. It consists of breaking down an image into the sensory input of features and objects that compose a scene description. This process is similar to semantic factoring in ontology (Kokla and Kavouras, 2001). It is a process of analysing and decomposing the categories of a given ontology into a set of fundamental categories. Thus, complex concepts are decomposed into simpler concepts out of which they are

constructed. Marr (1982) asserted that processing of visual information must begin with the perceived image of the real world, that is, the perceived image in the machine vision sense and not in the sense of mental imagery. Similarly, we can choose to dissect the geographical representation of the real world. Marr goes on to describe characteristics regarding the overall spatial arrangement of individual entities as primitive image elements. These account equally for spatial data representation, as the characteristics relate to the existence of surfaces that compose the world and create spatial configurations. These configurations take on a hierarchical organisation of spatial entities, which are often generated by a number of different processes, each operating at a different scale. Entities therefore have different levels of abstractions that subsume one another in hierarchical fashion. Items that are part of a configuration tend to be more similar to one another in spatial organisation, size and other attributes than to other items. In other words, similar objects have similar properties, forming groups or classes of objects. Furthermore, items in a configuration generated by a single process tend to exhibit some sort of organised pattern. These patterns result in a tendency toward smooth-shaped and non-abrupt boundaries between them.

As we learnt in chapter two, the inference of land use information from a topographic representation is a configuration problem. Configurational knowledge includes the ability to identify distributions, patterns, shapes, associations and relations of phenomena in both proximal and macro environments. It has been hypothesized that spatial knowledge requires only a declarative base and a set of procedural rules to allow understanding of complex spatial environments (Golledge, 1992). Most people have a common sense configurational understanding of spatial phenomena, as the questionnaire survey in chapter 3 confirmed. We learnt from the survey and related research that a lot can be discerned and named from topographic data. Although there is no explicit information about relations, such as buildings are near streets, or streets are in urban areas, etc., they are still observable in the data and reveal themselves to the user by visual inspection (Sester, 2000). Attributes that are typically stored in a database are material, type, and status. Others are hidden, in particular topological relations or proximities. These can be calculated, as GIS offers analytical tools to extract such information interactively. However, the higher-level semantics behind such relations and configuration is still missing. Semantic relations and intrinsic interrelations of the features themselves are often neglected (Tang *et al.*, 1996). That is why knowledge

about the objects and concepts to be found in the data has to be made available as a model of the relations between parcels and buildings, and other structures such as aggregated entities like urban areas.

The determination of geographical features is a complex process involving human perception and cognition. The analysis begins with the differentiation of data into categories, which must end with their reintegration into the whole image, as meaning is broken down into map elements and their structural entities that constitute the overall image. Shape, arrangement, similarity and proximity facilitate the making of vividly identified functional information within the map. There are clear interrelationships between map elements at different levels, as identified by Marr. We start with individual objects that form larger groups of objects based on similarity. These groups can be aggregated to blocks of objects, which are distinguished by their surrounding streets. These in turn form larger areas of homogeneity such as specific districts. Physical characteristics determine districts as thematic continuities which may consist of a variety of components, such as texture, space, form, details, building type, use, activity, etc. (Lynch, 1960). Homogeneities are characterised by similarities and their proximities such as direct neighbours and connectedness of groups, leading to a thematic unit. These elements operate together in a context providing a satisfying form to the observer. Most observers group the elements into larger organisations, or complexes that are sensed as a whole. In this holistic sense, people see the world as a whole rather than the sum of its parts. This means relations between a feature and the surrounding area are considered for the interpretation of the scene.

The preoccupation with parts, or map elements, rather than wholes is a necessary feature of an investigation into the interpretation of maps. To understand how the world is seen as a whole, we need to decompose objects into a set of sub-objects according to the user's interpretation of the reality. Geographic space can be defined as a finite, but not fixed, set of geographic entities having a recursive structure of partition and composition. A feature can both be composed of and be part of any other spatial or non-spatial objects. For example, a residential area is composed of houses, gardens and roads; likewise, a house is part of a residential area. The kinds of formalisms therefore needed for modelling geographic space in GIS, and particularly for semantic data modelling, are algebras or geometries dealing with such entities. Today GIS can

identify limited number of spatial relations like neighbourhood, containment or overlap relations, and only after very expensive, blind searches and computations, even to produce trivial results quite well presumable or known in advance. More complex, or conversely very intuitive, queries simply cannot be asked. What is needed are not formal models of spatial relationships among ideal polygons or lines, or topologic cells or simplices, but among houses and gardens, or streets and buildings (Nunes 1991). Further steps in the analytical direction require first of all what could be termed a taxonomic approach, that is, a systematic collection and specification of entities, their properties and relations. This will lead perhaps to a splitting of what is proposed now as geographic space in a number of interrelated sub-spaces, each one relevant to a kind of process. Second, a combinatorial approach determines how features aggregate to form a composite in each sub-space. Whether geographical entities will best be handled by set theory or not, this is in essence a configurational enterprise, not in purpose but in procedure. It is only after a successful differentiation and understanding of the kinds and parts before we can move on to consider the total system.

A conceptual model for exposing the link

A model is a simplification and abstraction of reality. It translates knowledge in a way that allows mechanical simulation. Modelling therefore requires that what is being modelled is made explicit. This includes the specification of things and relations between these things. Ontology provides the foundation for modelling. Since the model is an extract from reality, we need knowledge of the phenomenon under investigation. This includes the problem and its constituents, and a general understanding of how reality is composed, that is, what can be and how it is represented (Steimann and Nejdl, 1999). The starting point for an overall framework for accessing knowledge is therefore the conceptual model. A conceptual model is a general description of specific sets of entities and the relationships between these entities. The first step is to break down, or decompose, the thing to be represented into its elemental components, as outlined earlier. A geographic representation, in most general terms, is composed of entities, properties, and relationships. An entity in this context refers to spatial objects. Properties of objects are things that describe or characterise the object. Properties are crucial in explaining our ability to recognise and categorise things in the world around us: *"To categorise is to render discriminably different things equivalent, to group the objects and events and people around us into classes, and to respond to them in terms of*

their class membership rather than their uniqueness" (Bruner *et al.*, 1956, p.1). There are many different types of properties: Characterising properties (e.g. high), mass properties (e.g. cement, water), general properties (e.g. colour, shape), natural kind properties (e.g. river), artificial properties (e.g. road), and qualities (e.g. distance, area) as well as roles (e.g. land use) are considered as sorts of properties (Kokla and Kavouras, 2001). It is not enough to rely solely on mathematical descriptions of spatial properties and criteria since their definitions often fail to do justice to people's intuitive notions of what constitutes shape (Haggett and Chorley, 1969; Marshall, 2005). To capture the essence of human perception, understanding, reasoning and intuition, we need to model in a more natural way. For instance, on a set of spatial structures, such as topographic features, reasoning processes operate based on similarity judgement, proximities, shapes, sizes, etc., in ways similar to those investigated by Gestalt psychology (Thomson and Béra, 2007a). Concepts can be associated to these structures and processes to form a knowledge representation. Higher-level concepts can then be built incrementally from re-usable, primitive concepts as disposable one-time conceptual entities dependent on the information that is required by the user.

Since we are dealing with a domain that manifests itself with its underlying geographic existence, 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 super-ordinate levels of information. Geographic context includes information about geographic concept types, characteristics, relations and operations. These describe both the inner context, i.e., the context of features within a land use category, and outer context, i.e., the context among different types of land use categories. For the purpose of this thesis, we will only concentrate on the inner context to discern the composites of parts and features that make up the whole of a given land use category. Therefore, if we take the high-level concept 'residential', it can be related to the landscape through its make up of geographic objects, the objects' affordances and how they relate to one another to allow the use of that geographic space for human habitation.

In chapter 3, the questionnaire survey asked ordinary people how they related land use to the landscape character. The resulting 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. Table 6 shows some terms from people's conceptualisation of the residential land use. A person's conceptualisation is subjective, and involves knowledge, experience, perception and cognition. It happens on the mental level. On the other hand, its underlying geography represented through topographic data is an objectification of the real world. Its objects' inherent properties can be measured and made explicit, e.g. sizes, shapes, location. The representation is objective and happens on a digital, computerised level (Thomson and Béra, 2007b). The link between the two is already there: The mental conceptualisation refers directly to the topographic representation. It just needs to be exposed.

Table 6 Terms describing residential

Member categories	House	Block of flats	Front garden	Rear garden	Garden
Member objects	Building, low building	Tall building	Open space	Open space	Grass, open area
Purpose	Living accommodation, protection, residence, living area, housing, living in	Housing	Recreation	Recreation	Cultivation
Role	Living in, living, accommodation, provide comfort, shelter	Shelter	Entrance to house	Gardening	Enjoyment
Affordance	Sleeping, live, living in, living	Living	Walking	Relaxing	Sitting outside
Property	Small rectangle, square, individual rectangles, small,	Medium, rectangle	Very small rectangle	Large rectangle	rectangle
Taxonomy/ Partonomy	Kind of, part of	Part of	Kind of	Kind of	Part of
Topology	Meets/ adjacent to house, contained/ inside garden	Inside residential area	Meets house	Meets house	Adjacent to house

From the previous account, we can identify five important components for the conceptual model (figure 22): A collection of category types, a collection of relationships among these categories, a collection of functions (i.e., the purpose/role of categories), a collection of properties describing categories, and a collection of real world objects that are the constituents of the categories. Categories, or concepts, are the

contents of any representation. In figure 23, the category space describes the component concepts that make up the high-level category residential. These components consist of minimal meaningful units, such as primitive entities (building, roads, etc.), that combine to form higher-level meaningful composites of sets of elements (such as residential blocks and districts). Richer knowledge is derived from meaningful object configurations, special relations, perception (Gestalt principles) and context (Thomson and Béra, 2008). These relations form the relationship space, where categories are linked with one another. As identified by Barr and Barnsley (1997) land cover is organised spatially into discrete parcels whose morphological properties and the spatial relations between them convey information on land use and other higher-order 'meaning' about the scene. This meaning is described through the roles, purposes, or functions of the categories. The resulting function space asserts how categories are used, for example, a house provides accommodation. To be able to identify instances that belong to the categories, we require properties. The attribute space provides the necessary information to discern different types of categories. For instance, a detached house is typically larger than a terraced house. When we have to make a choice of which type of house we are dealing with, we just need to look at its properties and relations to other objects. Objects refer to the spatial features that represent these categories in the real world. For example, a house is a building. In a GIS, these features are represented as geometric objects with associated topological relations and classification attributes. Although GIS has been criticised for its reductionism because it artificially divides the world into parts, for our purpose this is particularly useful.

This vertical dimension of categorisation grounds the categories in the data by decomposing the rich knowledge to its finest level of detail in terms of its 'syntax'. This allows a one-to-one mapping between higher-level concepts and its representing geography (Thomson and Béra, 2008). I assume that words for categories, properties, relations, and functions, which make up the rich knowledge, are symbolic tokens for the things themselves (i.e., the real world objects). Therefore, the process of attaching meanings to words is essentially the same as attaching meaning to spatial entities. This process is termed semantic data processing (chapter 2). To adequately describe word meanings, we require morphological knowledge (forms of objects and their spatial order), syntactic knowledge (structural information and relationships), and semantic knowledge (meaning about context and how concepts relate to objects and relations in

the world). Syntactic knowledge is given by the syntax of objects. It relates to information about the structure of spatial objects (such as roads, buildings, land, and water), and how these comprise larger units that convey functional meaning.
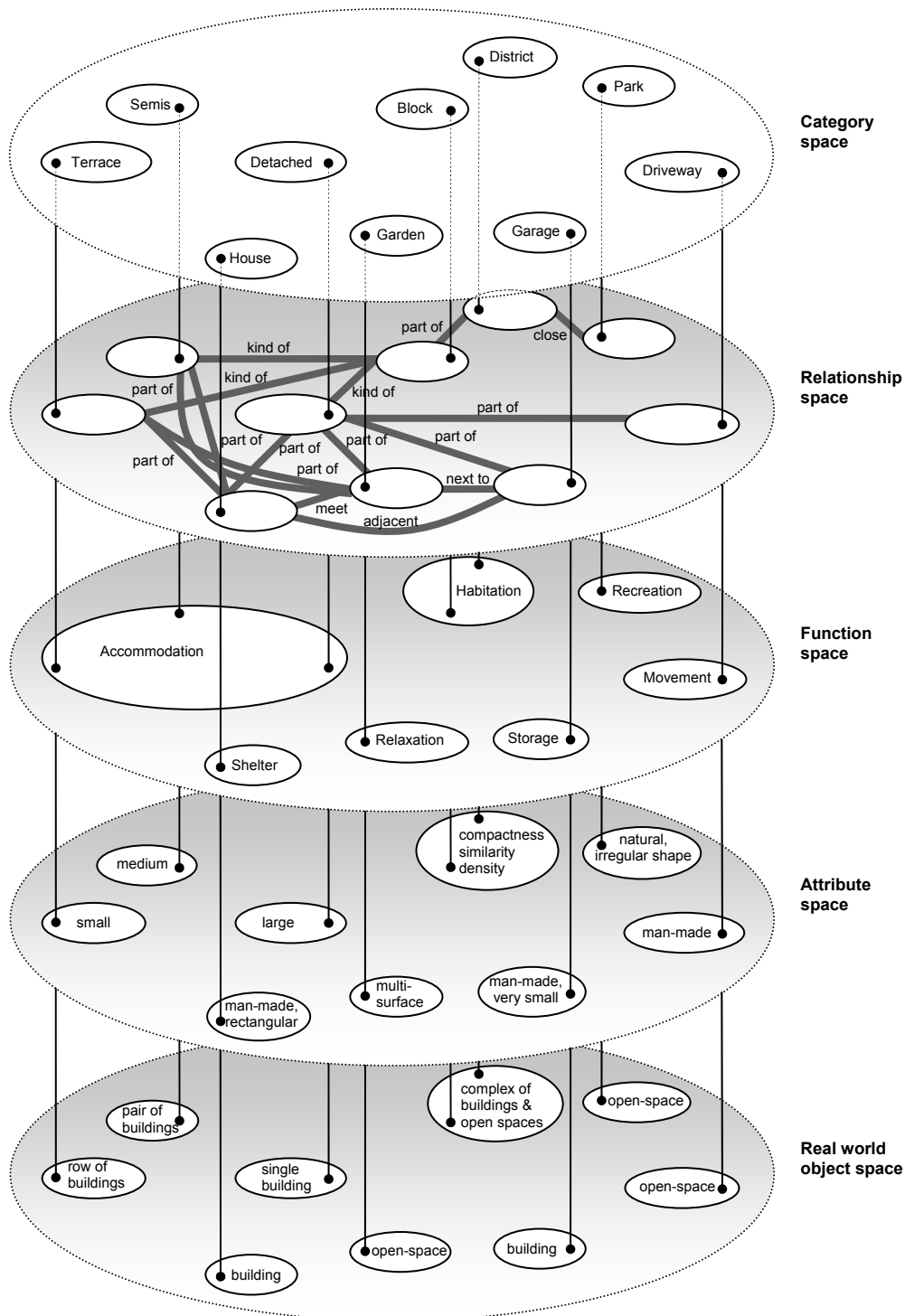
Figure 22 Land use conceptualisation of residential

The conceptual model relates to concepts that real-world objects possess, or at least apply to them in some way. These concepts in turn relate to the discrete data components within the geodata model where they are realised as features in a database.

There have been attempts to devise an ontology by reverse-engineering databases and vice versa (e.g. El-Ghalayini *et al.*, 2005; Astrova, 2004). However, the thesis focuses on common sense knowledge of what is needed from the data to construct higher-level knowledge through ontologies. Hart and Greenwood (2003), who built up detailed descriptions of individual real-world objects, realise the importance of the structural symmetry between concept, concept component, data component and feature in being able to make sense of multiple worldviews. This translates into four basic relations: An object has an attribute, an object belongs to a concept, a concept abstracts to an attribute, and a concept is a sub-concept of another concept (Hereth *et al.*, 2000). If we therefore constrain the concept and its attributes, then this has a direct effect on the object. This means the conceptual model is a set of goal-relevant constraints governing the representation of higher-level functional information. The ontology constrains the allowable relationships that may exist between concepts and their physical manifestation as data components.

These constraints are necessary to classify features stored in the data according to some high level concept. In traditional methods of classification, we first need to define the structures that we wish to recognise in the data. For example, Steiniger *et al.* (2008) define land use structures such as industrial and commercial area, inner city, urban area, suburban area and rural area for the automated compilation of medium scale maps. They assign measures to evaluate the structural properties including morphological (area, shape, corners, squareness, etc.) and relational measures (buffer, convex hull, etc.). Similar to Barr *et al.* (2004), they assess the separability of these structures through the given measures. Although concepts describe a very rich set of realities, we cannot live without measures when dealing with an objectification of reality. For example, certain conditions must hold for a given category, or concept, so that we can identify its members (Peuquet, 1988). Necessary conditions state that, for instance, for a house to be a house it needs to be a building. Graded conditions denote a central or threshold value for a property, such as the size of a building. A typicality condition is what we normally associate with a feature, such that a house is used typically for living accommodation. The problem is that we cannot get around the need for threshold values that determine certain properties, because we are dealing with concrete things or physical objects. In ontology, this aspect relates to concrete domains where quantities and measures need to be expressed with an otherwise purely terminological language. In

this context, we are dealing with a concrete domain of relations and structures in categorical maps. Although this allows us to make use of measures that are used for modelling cartographic relations both on the horizontal and vertical level (e.g. Neun and Steiniger, 2005; Steiniger and Weibel, 2005), we are faced with a major limitation. The way a person perceives the structure of a dataset cannot be reproduced by measures alone (Peter, 2001). Therefore, we are back to the initial problem: Whereas we can describe shape more intuitively through concept terms, the mapping of the concept to the physical object reduces it again to a mathematical definition.

The use of measures is unavoidable, but it can be seen as a means to translate human intuition into computational form. Since the reasoning still occurs at the conceptual level, the classification remains a human-centred process because the knowledge discovery process refers to the constitutive character of human interpretation (Hereth *et al.*, 2000). As the survey in chapter 3 reveals, respondents' success in interpreting plain topographic maps according to land uses is down to their ability to draw analogies from the known to the unknown. The cognitive use of a priori knowledge to interpret and categorise new observational data links this approach to the complex and interactive processes as led by human thought. The notion of schema relates to this process. A schema is 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 (Mennis *et al.*, 2000). The schema is not an exact representation, but is more like a general pattern. It provides 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. For example, 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. This could be that buildings will have a certain size, a certain shape, and a certain alignment. These criteria are determined by both knowing what an object is, and knowing where something is. The latter relates to the locational properties of objects such as containment, distance and direction. Kuhn (2004b), for example, poses the ontological question of the meaning of where and establishes the conceptual elements of a theory of location. The former relates to the detailed and precise

geometric properties, such as shape, relative orientation of component parts, size, and so forth. Downs and Stea (1977) refer to this type of knowledge as 'whatness' categories. It is the key to determine the identity and uniqueness of a category. Equivalent categories on the other hand are built around the similarities between places and objects commonly found in the spatial environment. They are classified and grouped together on the basis of shared characteristics. For example, land uses are unique. Residential can be clearly differentiated from recreational land use. However, blocks of flats, or districts of terraced and semi-detached houses share common characteristics because they all serve the purpose of residential accommodation.

To translate this ability into mechanised ways, we require a priori knowledge of a general representation of the domain of interest. Torres *et al.* (2006), for instance, successfully use a priori knowledge of a satellite image to aid the supervised clustering by adding its intrinsic semantics. The conceptual framework described here provides the means to define and formalise higher-level knowledge as an ontology that maps directly onto the topographic data to interpret its immanent functional meanings. Figure 23 illustrates the proposed methodology. Domain concepts of higher-level semantic information are derived from the external schema, the users. Data concepts describe the underlying geographic data making its primitives explicit (discrete, identifiable entities with a geometrical representation and descriptive attributes), i.e., the internal schema. By combining these two approaches into a shared conceptualisation of the application ontology, that is, into the conceptual schema, it becomes 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. The classification is therefore constrained by the ontology's specification. Features that meet these specifications become instances of the high-level concept and inherit the defined higher-level semantic information. For example, having various levels of semantic definitions for the category residential allows the expression of this category at a high level as districts and blocks, and at finer levels as terraced, semi-detached and detached houses, and row of buildings, pair of buildings, and single buildings, respectively. The high-level semantic description residential can then be inherited down to the finest level, i.e., that of individual objects that meet the schema's criteria of a certain size, shape, type, etc.
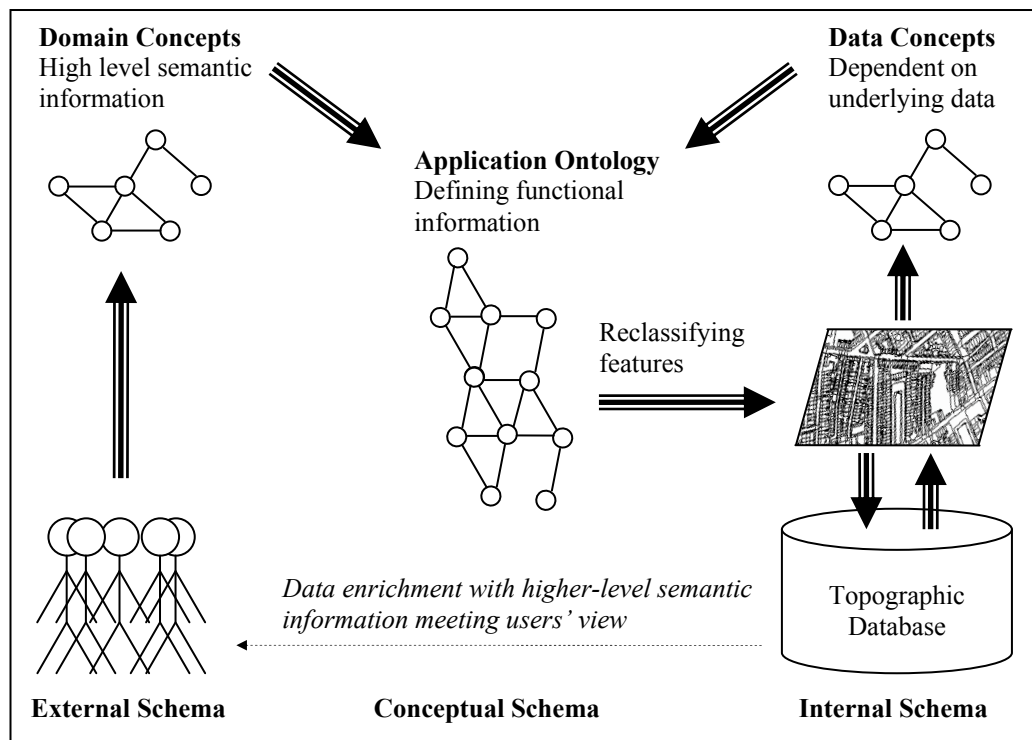
Figure 23 Proposed methodology

The advantage of using ontologies is that they give a concise, uniform, and declarative description of semantic information independent of the underlying syntactic representation of information bases (Kashyap, 1999). This means when changes occur to the underlying data, this does not affect the defined high-level knowledge. With an ontology we can reclassify the data's topographic instances and derive a set of areas depicting the human activity that takes place on that geographic space. 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 directly integrating ontologies into the conceptual representation of designing and implementing the physical database design (Fonseca *et al.*, 2003). Although it is not a new endeavour to incorporate cognitive principles into geographical databases and to derive a unifying, semantic data model (e.g. Peuquet, 1988; Mennis *et al.*, 2000; Mennis, 2003), often research remains at the conceptual level. Despite much research on the use of ontologies in the geospatial domain, authors have often either left the relationship to the data model undefined, or have tied it to one physical implementation method (Hart and Greenwood, 2003).

The ontological approach to map interpretation

Interpretation is a fundamental human cognitive ability. It is a knowledge intensive process that is decisively shaped by the way common-sense knowledge and experiences are brought to bear. We have seen that interpretation is the ability to assign meanings to input data through the assignment of one or more concepts or categories. We have also seen that interpretation can be qualitative, i.e., assigning a qualitative concept or category, or quantitative, i.e., assigning a numerical value, or measurement. Brey (2005) argues that computers extend the memory, interpretation, search, pattern matching and higher-order cognitive abilities of human beings by performing cognitive tasks autonomously. Although computers are systems in which symbol structures are capable of representing objects in the real world that are manipulated in intelligent ways, they can only work by reducing them to information-processing tasks. So far, I have considered ontology as a framework to represent information to describe a certain domain of entities that are related to each other with a particular notion of reduction to find a simple and systematic theory. Ontologies, however, are also more easily accessible to automated information processing.

The previously described conceptual model illustrates how implicitly represented knowledge within a spatial database, such as residential area, can be inferred from knowledge that is explicitly defined through the different ontology levels described in chapter 4. Figure 24 shows how the bottom-up and top-down approaches of the modelling come together. By going bottom-up, we adopt an agglomerative approach where each primitive element, describing a separate data entity (e.g. building) becomes part of its parent aggregate concept (e.g. terraces) until the high level interpretation (i.e., residential area) of the scene is reached. The top-down approach defines the part-whole aggregation based on certain criteria and rules from a high-level conceptual view, which will constrain the reasoning. The low-level concepts describe the entities stored in the database, and can be seen as a surrogate representation. The link between the conceptual model and the data level is achieved by having an application classify the data's instances in terms of the semantic categories that the conceptual model provides, starting with the low-level ones and incrementally building up the high-level interpretation. As a result, the interpretation process needs to be modelled as an incremental construction process with the goal to create and verify any instances that may be useful for the overall map interpretation.
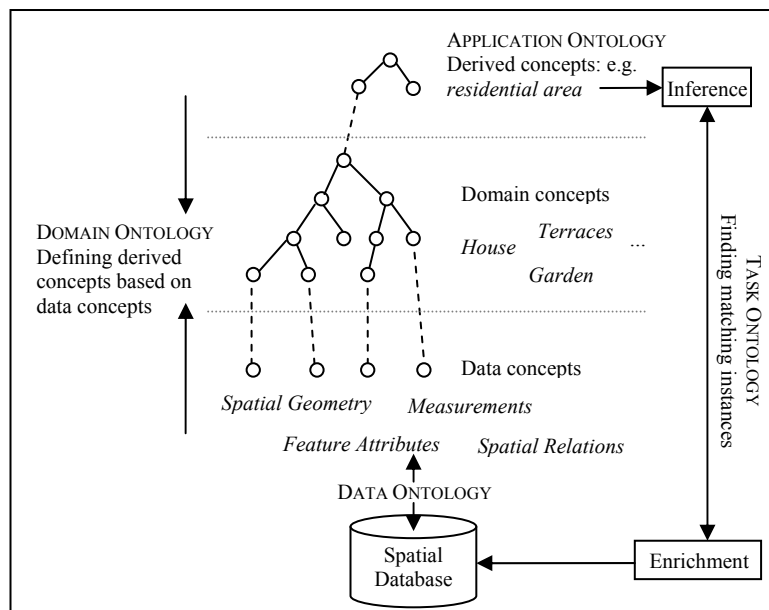
Figure 24 Relating higher-level concepts to the data structure

This methodology uses description logics in a way similar to Neumann and Möller (2008) for scene interpretation, where the recognition of the whole (scene or map) arises from the recognition of its parts (aggregate concepts). It follows the tendency to continually abstract through simplification, as people perceive space as a composition of simple geometries and similarities (Batty and Longley, 1994). To create a diversity of patterns and parts through a system, generic design guidance is required which specifies the elemental units or sets of units to be recognised dependent on the purpose of the classification. Form plays an important role, since space is not only observed and understood in terms of its spatial pattern, but is composed of such elements. Modelling higher-level meanings is difficult because of the mix of heterogeneous activities and uses, which have a high complexity, threatening the classification of their geometry, as well as impeding objective and consistent categorisations (Batty and Longley, 1994). Nevertheless, through simplification a system structure can be built composed of elements and relations that decompose into further subsystems arranged into distinct hierarchies of taxonomies and partonomies. This structure enables inference of higher-level information based on reasoning about its defined concepts and finding relevant instances. Although we have to be careful not to force the diversity of functions into a narrow range of concepts, a parser working according to some configurational rule can incrementally build up the semantic interpretation of a map scene using the corresponding object semantic rules of spatial composition.

The recognition is simplified to the detection of objects that must have a specific spatial configuration (Haarslev *et al.*, 1994). The inference is determined deductively from the primitive characteristics of objects visible in the map. The key notion is to provide a conceptual description of the complex structure in the map we wish to recognise, such as residential area, and to find its instances that are to be classified accordingly. Given appropriate high-level knowledge structures, far-reaching interpretations may be obtained including propositions about parts of the maps for which there is no direct evidence at all. Hence, higher-level knowledge can be inferred and assigned to map primitives, and with that we can instantiate land use information and assign it to constituent land cover parcels. This is made possible because land use is treated as a configuration problem (chapter 2), which provides the foundation for logical scene interpretation as applied by vision recognition systems in artificial intelligence (e.g. Schröder, 1999; Möller *et al.*, 1999; Neumann and Möller, 2008). The approach is equivalent to logical model construction (Hotz and Neumann, 2005). For example, configuration systems have been developed in support of tasks where parts (usually technical components) have to be configured to form a system that meets the given specification (see the Lego example in chapter 2). Here the parts are land parcels of a topographic map. Therefore, scene or map interpretation formulates as a finite model construction task that is implemented through constraint satisfaction.

Model construction in this sense applies to the symbolic description consistent with conceptual knowledge about the world and concrete knowledge about the scene. The interpretation is formalised as a concrete application of the axiomatisation of the required knowledge through first-order logic to provide a formal definition for the expected result of the interpretation problem (Schröder, 1999). This means that a formal mapping is constructed from constant symbols and predicates of a symbolic language into the corresponding entities of a domain such that all predicates become true (chapter 6). Hence, a valid map interpretation must be a model of the conceptual knowledge and the map data, as it connects constant, predicate and function symbols of the high-level domain with corresponding individuals, predicates and functions of the represented data domain. Thereby, the mapping between the two domains becomes a model if it causes all symbolic expressions of the conceptual knowledge and the map-specific knowledge to become true. In that respect, an interpretation of a map is a partial description in

terms of instances of concepts of the conceptual knowledge base. It is partial because only parts of the map and a subset of the concepts are interesting in general, depending on the pragmatic context (Neumann and Möller, 2008). Consequently, the interpretation process needs to be modelled as an incremental construction process with the goal to create and verify any instance that may be useful for the overall inference.

Figures 25 illustrates the basic idea of how a formal interpretation based on ontologies compares with the human interpretation. To make new knowledge accessible within the data, we require the two inputs described above. Firstly, the knowledge base consists of evidence from the map in the form of its low-level features. The so-called A-Box of a description logic system stores all the asserted information about the facts, similar to the information we capture with our eyes when reading a map. The retinal image therefore equates to the described knowledge in the knowledge base. Secondly, the so-called T-Box of the knowledge-based system captures the logical assertions on concepts in the ontology according to the described conceptual model. The ontology therefore provides the necessary a priori knowledge of spatial objects, their properties and relations to one another in the map scene akin to a human-constructed physical world.

Both concepts and facts are described using a highly expressive object description language, and are embedded in a taxonomical hierarchy. A compositional hierarchy is induced by the special structural relation part-of. Constraints among concepts pertain to properties that are in turn specified by parameters with restricted value ranges or sets of values (Hotz and Neumann, 2005). The interpretation process is to instantiate concepts of the knowledge base by checking for which concepts the relevant attributes and constraints in the map are fulfilled. This means individuals in the A-Box must interface to the data level and map onto the high-level concepts in the T-Box. Only then, a map description, for example that of residential area, can be generated and its symbolic meanings, i.e., semantics, can be assigned to the map primitives. We speak of ontology-assigned meaning.
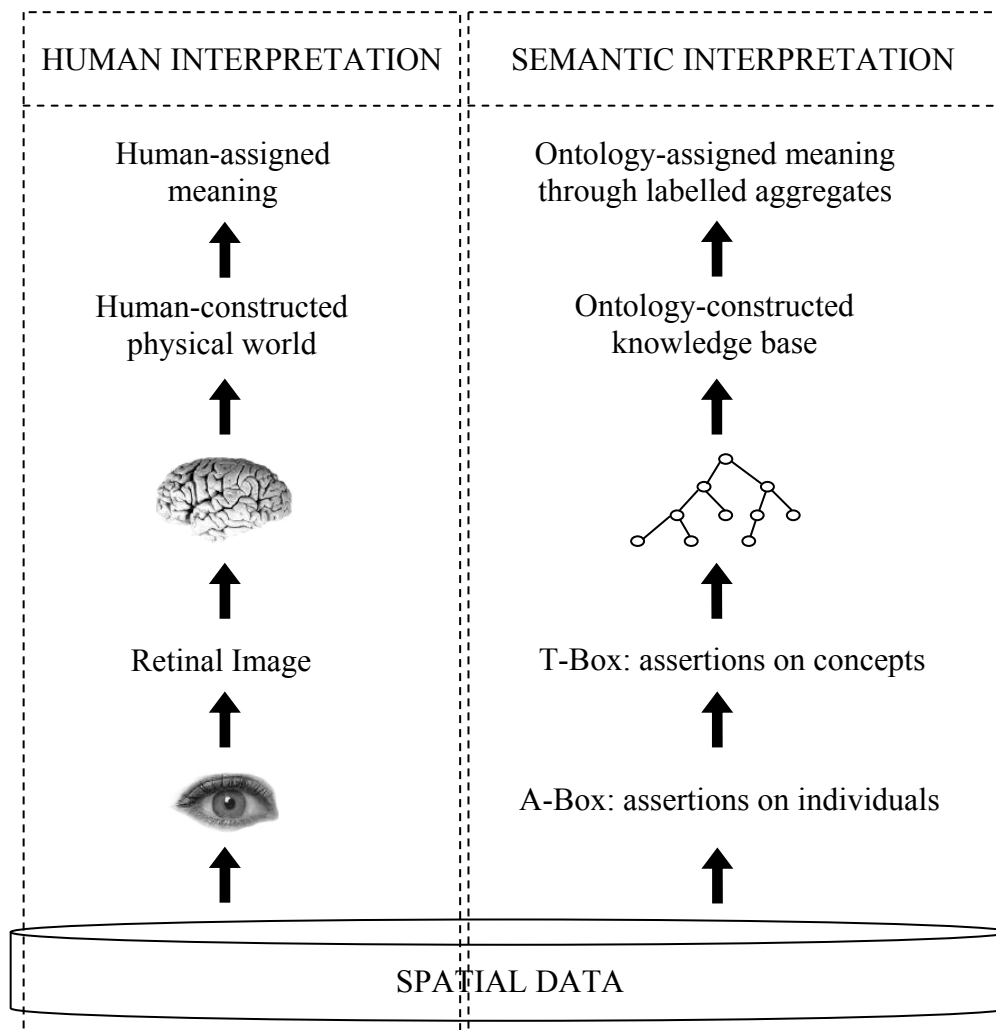
Figure 25 The ontological approach to map interpretation

The advantage of such a conceptual model is that it always can be further extended by matching a new description or concept into the taxonomy of existing concepts and linking it directly to its most specific subsumers and the most general concepts that it in turn subsumes. Plus, rather than relying on surface based comparisons as in traditional classification algorithm, information retrieval based on the semantics of the data model provide large potential for information access (Bresciani and Franconi, 1996). Different cognitive situations can be treated with interpretation strategies based on the same conceptual basis. This feature distinguishes this approach from rule-based or other deduction-based approaches where interpretation strategies are much narrower defined. Mostly two-level representations have been proposed to integrate logical representations for qualitative spatial relations and quantitative information (Haarslev *et al.*, 1994). Next, we will learn how semantic aggregation at multiple levels of abstraction provides the necessary framework for inferring high-level representations such as the residential district or area.

5.2 Semantic aggregation

For urban planners, the pattern specification of urban areas is important to develop a meaningful and manageable system of description, addressing physical patterns such as urban forms and networks, and relationships between different patterns, at different scales. Such a specification assists in interpreting physical and spatial patterns existing on the ground or modelled (Marshall, 2005). Their discrimination provides useful information. The town planner, for instance, might be interested in the distribution of neighbourhoods, their facilities, and green spaces. Because the visible world is viewed at multiple scales and degrees of detail (chapter 3), there can be no single correct or definitive way of classifying patterns. However, a diversity of overlapping pattern types and themes are both appropriate and inevitable. Often a balance needs to be struck between having too few broad categories or too many narrow ones (chapter 2). Taken to the extreme, we may end up with a single category into which all actual patterns are lumped. The proposed system in this thesis provides a dynamic solution by accommodating different levels of patterns. Depending on the requirements, you can have very fine levelled categories, such as terraced, semi-detached, and detached house patterns, and a very broad, overall category of residential area. This avoids over-generalisation or stereotyping. The conceptual model orders the different kinds of categories of pattern classification (i.e., configurations), and harmonises the use of language and terms by explicating the typologies' semantics.

To provide improved information access, we need to organise data into intelligible and easily accessible structures at various levels of detail (Bresciani and Fanconi, 1996). Over the course of this thesis so far, we have learnt that map reading is a process that accesses the spatial knowledge stored in the 'cognitive map'. According to Kuipers (1978), it is essential to have a theory of the representations in the cognitive map to make computational theories of these processes. It is typical for high-level abstractions to resort to common-sense knowledge, beyond the knowledge about geographic phenomena. Land use, for instance, describes a map at an abstraction level above the single-object trajectories present in spatial data requiring qualitative and symbolic representation. Therefore, functional information is typically embedded in a compositional hierarchy with increasing abstraction towards higher-levels. Spatial relations between constituent parts and objects, as well as their morphological

properties define functions. These relations must be evaluated efficiently to support the recognition of functions in the spatial data. The recognition then requires incremental part-whole reasoning on the different levels of abstraction. Consequently, the conceptual model builds up descriptive primitives (from the most detailed level, i.e., the data), which will become successive groupings producing hierarchies of entities and spatial patterns.

We have seen that pure mathematical definitions fail to do justice to our intuitive notions of what constitutes shape. Especially when we attempt to generalise shapes into high-order features, we must also abstract their conceptualisations. The first abstraction involves mapping identified features into geometric shapes and symbols with sizes, styles and possibly colours. This produces a useful description of the scene depicted in the topographic data, whose initial representation is a collection of polygons, lines, and points stored in a spatial database. The next abstraction involves simplifying, culling, and coalescing graphic objects into smaller versions of the things they represent. In fact, much of the generalisation literature is stuck at this level (e.g. Duchêne *et al.*, 2003; Christophe and Ruas, 2002; Gaffuri and Trévisan, 2004; Li *et al.*, 2004). The third abstraction is to generalise the concepts that the symbols represent, such as residential area, and then depict those. Usually domain-independent algorithms (from generalisation) extract, characterise and label the components, and then deliver more generalised representations. Here is where ontologies can make a difference (e.g. Camacho-Hübner and Golay, 2007). One way to enrich cartographic data without completely recompiling them is to build associated ontologies conveying the meaning of data items, their properties, and relationships for operational purposes (Dutton and Edwardes, 2006). Initially, the mapmaker sets down spatial semiotics to depict the real world with a host of tacit semantics based on mapping standards and lack of thought. Often ignored is therefore the role that perception has in specifying abstractions of reality, that is, how you generalise maps depends on how you perceive the world and how you specify it.

Cognitive psychology in particular supports the idea of hierarchical representations (Peuquet, 1988), because organising pieces of information into larger, meaningful units is a universal cognitive principle. Grouping is one of the effects resulting from this principle. The grouping mechanisms are connected on a functional level by their spatial

grounding and thus contribute to cognitive control during spatial reasoning (Engel *et al.*, 2005). This means, groupings are not only reflected in the various representations, they also entail one another on the basis of a common principle. However, the determination of meaningful groupings can almost never be achieved directly from the scene (i.e., the observed data), but from pre-existing knowledge concerning the nature of the given phenomena involved in the scene. In terms of our conceptual model, this means that although a map can be seen as a series of levels at different scales (Chaudhry, 2008) – that is, from an image at street level of individual buildings to levels of neighbourhoods and districts – we need some pre-existing knowledge concerning the nature of the given phenomena involved. With such knowledge, we can then determine the structure of each level and its subordinate elements. In this fashion, it is possible to model the knowledge of each stage and reason about it. This incrementally builds an interpretation from the individual object, say house, to its high-level, superordinate category, say district. Consequently, the gap between the data and the conceptual model is bridged via a range of representations, which connect the input data to the output interpretation (Sowmya and Trinder, 2000).

The purpose of abstraction is to provide richer and more expressive concepts with which to capture more meanings than were possible with classical data models (Tang *et al.*, 1996). Hierarchical levels of geographical features are useful for expressing these abstractions. Whereas semantic granularity addresses the different levels of specification of an entity in the real world, spatial granularity deals with the different levels of spatial resolution or representation at different scales. A combination of five abstraction mechanisms including classification, generalisation, specialisation, aggregation, and association, make these different levels of abstraction possible. Classification is the mapping of objects that share the same behaviour or characteristic into a common class. Generalisation is the mechanism to form a general super class by combining several classes of objects of similar type or with common properties and functions. Specialisation creates specialised classes that inherit structure from higher-order object classes. Aggregation is the collection of a set of subclass objects, each with its own functionality to form a semantically higher-level parent object. All these mechanisms operate on the 'is-a' and 'part-of' relations, which determine the hierarchical and combinatorial arrangements of classes.

From an ontological perspective, these kinds of processes and relations map easily to semantic relationships between two or more concepts in an ontology (Fonseca *et al.*, 2003). The relations map to the notions of hypernymy, hyponymy, mereonymy and synonymy as applied to ontologies (Kokla and Kavouras, 2001). Hypernymy and hyponymy are semantic relations defined between words and word senses. Hyponymy, subtype/ supertype, or a-kind-of relation is the subordination/ superordination relation defining the taxonomy of concepts. The hyponym inherits all the characteristics of the more generic concept and adds at least one characteristic that distinguishes it from other hyponyms. Meronymy/ holonymy denote the part-whole relation. Synonymy refers to similarity in meaning, as we analysed in chapter 3 for terms describing role, purpose and function of a given concept. These relations are the operating factors that form higher-level classes of objects from other classes, creating different abstractions and inheriting from their superordinate classes. This means some objects can be defined entirely from other objects. Mennis *et al.* (2000), for instance, differentiate between objects that are derived directly from their observation in the data, termed atomic, and those objects that are composed solely of other objects, termed composite.

When dealing with spatial data, additional relations come into play. According to Fonseca *et al.* (2003), topologic relations are fundamentally important to the definition of spatial integrity rules, which in turn determine the geometric behaviour of objects. Other relations that describe geographical phenomena include direction, distance, nearest neighbour, adjacency and containment, which associate different locations over a single, continuous surface. The spatial as well as membership relation among various entities determine the entities' positions across a given, common level within the hierarchy. Taxonomic and aggregation relations, on the other hand, allow association among entities at different levels up and down the hierarchy. All these relations are essential for the interpretation steps. Aggregation in this sense relates to the act of inferring an aggregate from parts (part-whole reasoning). Specialisation means tightening properties and constraints, either along the taxonomical hierarchy or by checking objects for possible roles in aggregates. And generalisation is the step of instantiating the parts of an aggregate if the aggregate itself is already instantiated.

The primary element of the conceptual model is the specification of how the elements of the representation are defined, combined and added as new information. A graphical

understanding of the discussion of this approach is illustrated in figure 27. The functional reality can be grasped in adequate fashion only by taking entities at a multiplicity of different granularities into account. According to Batty *et al.* (2003) whether or not a place has more than one function, land use or activity, depends on the size of that place. For example, in our abstraction hierarchy (figure 26), at the low-level we usually deal with a single activity. At higher-levels, the abstraction encompasses a larger area, such as blocks, block groups or districts, where more than one function can take place. For instance, when we aggregate everything to one space – the level of the residential district, say – then multiple activities may take place there, such as recreation if a park is contained within the district.

The theoretical attention to the representation of the spatial or formal system gives rise to a whole family of representations of the same spatial system, each one relevant to some aspect of its functioning (Hillier, 1996). It is therefore normal to combine representations, literally by laying one representation on top of the other and treating the resulting connections as real connections in the system. This is a typical procedure in the generalisation literature, where urban structures are built based on their component patterns and inner organisations (e.g. Gaffuri and Trévisan, 2004; Boffet and Serra, 2000; Boffet, 2000; Larive *et al.*, 2005; Ruas, 2000). For example, a town is composed of urban districts, which in turn are a collection of urban blocks. The block is shaped by and consists of building groups. Further, the group is determined by the building alignment. These organisations link to the micro, meso and macro levels of generalisation. However, the structural consideration alone here is not sufficient. Just as Hillier believes that taking pairs or even triples of representations together yield formally or functionally informative results, the model represents space in terms of the type of function in which we are interested.

The representational units for modelling this incremental construction process are aggregate concepts. An aggregate specifies a set of objects with certain properties and relations that together constitute a meaningful scene entity (Hartz and Neumann, 2007). A configuration typically constitutes a complete model for an aggregate specified by the configuration task. Hence, an aggregate represents object configurations and other high-level structures that form part of the compositional and taxonomical hierarchies imposing the main structures of the high-level conceptual knowledge base.

Area Aggregates:
A definite area of a place, distinguished from adjacent regions by some special quality or function, reflecting a primary character such as residential (inner city, suburbs, rural).

Residential Area (suburbs)

District Aggregates:
Any tract of land, usually of vaguely defined limits having some common characteristics within which one or more principal land uses are permitted along with their accessory uses (e.g. park).

District-semis (dominant)   District-detached (dominant)   District-terraces (dominant)

Block Aggregates:
Plots of land surrounded by public streets.

Block-terraces   Block-semis   Block-detached   Block-mixed

Housing Aggregates:
A composition of primitives considered as a basis of formation for the concept 'residential'.

Terraces   Semi-detached   Detached

Primitive Aggregates:
Concepts without parts.

House   Garden   Out-building   Road

Figure 26 Semantic aggregation

The main motivating criterion for defining an aggregate is to provide a coherent description of entities that tend to co-occur in a map scene (this is regardless whether information is explicitly or implicitly contained in the map). An aggregate expresses the

properties and constraints that make an object, or a particular set of objects worth being recognised as a whole. It may consist of other aggregates depending on its constraints that relate constituents of different parts to each other. For example, the aggregate concept semi-detached house consists of the concepts house, garden, and optionally an outbuilding. This induces a compositional hierarchy that is built on top of primitive entities contained in the map as illustrated in figure 26. The meaning of aggregates evolves throughout the hierarchy by developing a significantly different definition at high-level from the definitions of the features' low-level classification. The resulting abstraction hierarchy describes how the different definitions of spatial objects link at several scale levels.

This abstraction hierarchy has its roots in models of perception. For instance, if I see something for the first time at some distance, I may not see all the details of the object nor may I be able to consider all possible instances of details that I am missing. Nonetheless, I will be able to make certain observations and inferences about the object. This is possible because many observations and inferences are independent of the missing details. Knowledge about these missing details would allow for a refinement but not require corrections of the observations. Consequently, the bottom-up view suggest that incomplete knowledge is due to omission of specifications, thus it can be completed by considering the set of possible augmentations. The top-down view suggests that incomplete knowledge is due to possible distinctions of details which are not made, thus by ignoring details, we can deal with coarse knowledge. With this view of the world, we organise knowledge hierarchically according to the level of detail that is available: Coarse knowledge corresponds to the higher levels, and detailed knowledge to the lower levels in this organisation. On any level certain inferences can be drawn. These inferences can be expressed in terms of knowledge represented on the same level of detail or of a higher or lower level. One advantage of this approach is that inferences that can be drawn on a higher level subsume several corresponding inferences on lower levels. If additional knowledge about details becomes available, the inferences are refined (Freksa, 1991).

What we want is to represent knowledge in such a way that few concepts capture rich situations. This means that gradual transitions from one concept to another should be captured by introducing only the number of intermediate concepts required to achieve

the goal. As suggested by Freksa and Barkowsky (1995), the model takes into account the neighbourhood structure of geographic entities according to a physical model, the vertical and horizontal neighbourhood structure of the spatial concepts, and the correspondence between concepts and geographic entities. The problem is to understand how levels are constructed, i.e., to know the criteria used in the selection of particular aspects of a situation and how these perspectives and aspects are combined into a representation that can be used to derive the higher-level abstraction. The number and contents of the levels and their relations characterise a particular modelling domain or even a specific problem (e.g. other function categories), and thus cannot be defined in general. The hierarchy needs to be adapted for different purposes.

Conclusions

The need to develop a tool for image interpretation that segments a map and automatically associates geometric regions on a map with semantic labels has been long recognised (e.g. Esposito *et al.*, 1997). In fact, the concepts and methods people use to infer information about geographic space become increasingly important for the interaction between users and computerised GIS. This chapter explained how these methods can be emulated by producing a semantic abstraction hierarchy that links high-level concepts to the data. Thereby the question was addressed of how to bridge the gap between knowledge, i.e., conceptualisation, and geographic data, i.e., representation. Once knowledge is available in machine interpretable form via ontologies, it can be linked to a topographic database through a conceptual hierarchy creating a mapping between high-level knowledge structures and those present in the data. Since space can be observed from many different viewpoints and at different resolutions, the same objects may be represented differently, depending on the purpose, indicating the point of view to take and the level of detail to be included. This is generally termed multi-representation of spatial objects (Timpf and Frank, 1997). The representation of the spatial objects at different levels of resolution leads to a hierarchical representation where more and more details are included as one descends the hierarchy. Here, the interest lies with accessing high-level knowledge by ascending the hierarchy.

Categorical database generalisation, for instance, relies on the exploitation of hierarchies that are inherent to spatial data (Liu *et al.*, 2003). Timpf and Frank (1997)

have also identified the benefits of hierarchical reasoning, because it uses the level of detail appropriate for the task. We humans constantly isolate the relevant aspects and relate them to one another to achieve intellectual efficiency. This efficiency is necessary for successfully operating in the world (Freksa and Barkowsky, 1995). Spatial hierarchical reasoning denotes the deduction of information from a representation of a spatial situation. It also applies the economic principle that a task is solved with the least amount of effort. Therefore, spatial hierarchical reasoning requires a method to derive less detailed representation from the most detailed one. This is achieved by modelling incremental structures along the hierarchy with different levels of semantic abstraction.

When people deal with spatial objects, they usually automatically attach higher-order meaning. The processing of spatial information therefore begins with the perceived image of the real world. Once we select directly observed phenomena and abstract them into key characteristics of the scene or map, we then interpret these characteristics using pre-existing knowledge. For example, directly observable features of a topographic map are buildings, roads and open spaces. We can interpret these structures as urban areas, inner city, or residential depending on their configurational properties. Whether this information is purely conceptual consisting of high-level abstract objects, or whether this information is purely representational as in a map, currently there are no models for a comprehensive treatment of different kinds of spatial concepts and their combinations that are cognitively sound and plausible.

Modelling is not a straightforward task. The difficulty of visual object recognition in general – let alone the interpretation of information – is the complexity of the scene and the generality of the object models. The adoption of a model-based recognition method based on a hierarchical organisation means that we are faced with the limitations of any kind of tree structure for representing knowledge. Indeed, a tree can represent many real-world phenomena, as we learnt in chapter 2. Whether it is correct to do so is a separate issue (Hirtle, 1995). Nevertheless, adopting a knowledge representation framework, namely, providing sufficient representations to allow reasoning about geographical situations and land use phenomena, means that a model has both psychological and epistemological relevance. For example, the model provides a psychologically relevant representation of geographical features because of its link to

human conceptualisation and the way its abstractions are partitioned according to Gestalt principles, that is, the way people interpret topographic maps. Because higher levels of abstraction are less detailed than lower ones, moving up one or more levels allows one to see the forest through the trees. This view is characterised by the use of conceptual categorisation. One convenient categorising process for spatial data involves partitioning space into recognisable and acceptable units such as country, region, community, neighbourhood, and so on (Gale and Golledge, 1982). In the geographical context, it is therefore possible to identify a hierarchy of scales that people use to partition space.

The conceptual framework is not directly concerned with the physical transformation of objects into higher-level representations (e.g. Liu *et al.*, 2003). Rather it describes semantic generalisation of how existing instances in a database can be thematically enriched and therefore accessed in more meaningful ways. Subsequently, the enriched information can then aid generalisation algorithms to derive the necessary physical representations. Indeed, there are algorithms to perform the same steps of deriving building alignments, blocks of buildings and whole districts (e.g. Regnauld, 2001). These algorithms perhaps provide fast computation, but they remain a black box method with the essential semantics hidden behind the algorithms. The benefit of adapting the knowledge representation paradigm is therefore the explicit modelling of knowledge. The domain of geospatial concepts is separated from the domain of data representation according to the layered ontology architecture of the conceptual framework. This separation ensures that the conceptual model is not affected when changes occur to the underlying data representation. Furthermore, the model can be easily extended by including concepts in its taxonomic and partonomic hierarchies. This means that inferences carried out on the basis of coarser knowledge should remain valid when additional knowledge becomes available. However, knowledge formalisms still put us at the mercy of mathematical theories such as sets and logic. It follows that because spatial data are a concrete domain of physical and quantitative nature, we cannot divorce ourselves entirely from the mathematical definition necessary to constrain properties.

Nevertheless, the model allows incorporation of complexities of spatial data and relations in the database, thus capturing higher-level meanings. Its abstraction hierarchy

encapsulates data and functions in user-defined object classes allowing holistic representation of features. User-defined means that the concepts are designed to represent characteristics of features that are relevant to a particular application, say deriving residential areas. Geographical features are identified according to their common characteristics and function, and are grouped into higher, more meaningful configurations. For example, if instances meet the constraints or rules of a given concept definition, then they inherit that corresponding higher-level concept. Inheritance of properties and structure from superclasses to subclasses makes abstraction of geographical features possible. Aggregating objects is therefore a powerful means to achieve abstraction in spatial representation. Semantic aggregation enhances the representation of features in a holistic way because each feature can describe the total information about a location and the relations with other features for a specific application. The use of ontologies for designing and representing these aggregate concepts makes the more meaningful abstractions and their associated representations possible and practical. The result is a more complete digital representation and spatial description of geographical phenomena.

The conceptual framework presented here at a general, descriptive level will easily map into expressive description logic, as I will explain in the next chapter. I will describe the model in its formal, implementable form. Further, I will give concrete examples of how to infer an interpretation of a map that is consistent with conceptual knowledge, evidence, and context information. The link between the data model and the conceptual model is achieved by having an application (e.g. Protégé) classify the data's knowledge in terms of the general semantic categories that the conceptual model asserts. There is a clear partition between the conceptual model described in this chapter, its formalised form (chapter 6), and its application with off-the-shelf knowledge-based software (chapter 7). Hence, this thesis goes beyond pure conceptual work (e.g. Mennis *et al.*, 2000; Peuquet, 1988) by considering how objects and classes are actually generated from observational data.

Chapter 6

Formalising a Representational System

"The only way to rectify our reasonings is to make them as tangible as those of the Mathematicians, so that we can find our error at a glance, and when there are disputes among persons, we can simply say: 'Let us calculate, without further ado, to see who is right.'"

– Gottfried Leibniz, 1685

Knowledge representation refers to the general topic of how information can be appropriately encoded and used in computational models of cognition (Wilson and Keil, 1999). The practical goal of constructing frameworks for knowledge is to allow computational systems to attack knowledge-intensive problems such as real-world reasoning. Ontology only forms a part of such a system. It describes a vocabulary for talking about a domain. In knowledge bases, there is a clear distinction between terminological and assertional knowledge about individuals and their membership to concepts and roles described by the ontology (Schaerf, 1994). This knowledge is needed to solve a problem to answer arbitrary queries about a domain. Therefore, a representational system that formalises knowledge must contain what the represented world is, for example real world entities or spatial data in this case, and what the representing world is, that is, the conceptual model. It must know what aspects of the represented world are being modelled, what aspects of the representing world are doing the modelling, and what the correspondence between to two worlds is (Palmer, 1978). In other words, the nature of representation is the existing correspondence, or mapping, between objects in the represented world (evidence) and objects in the representing world (conceptual model), such that at least some relations in the represented world are structurally preserved in the representing world.

Consequently, the foremost role of knowledge representation is to substitute for the things in the world to enable a machine to determine consequences by reasoning about the world rather than taking action in it (Davis *et al.*, 1993). Reasoning is a process that goes on internally, in our minds for example, but most things it wants to reason about

exist only externally. The coupling of high-level knowledge with a GIS also requires two representations to be linked: Spatial data in a GIS with application specific knowledge (Miller, 1994). This unavoidable dichotomy is a fundamental rationale and role for knowledge representation. It functions as a surrogate inside the reasoner by mapping between real world representations and conceptualisations of a given domain. The correspondence between the surrogate and its intended referent in the world is the semantics for the representation.

Because knowledge representation is a surrogate, it is unavoidably a set of decisions about how and what to see in the world. Selecting a representation thus means making a set of ontological commitments in terms of the concepts, properties and relations that represent relevant knowledge. This is necessary because representations are imperfect and the complexity of the natural world is overwhelming, forcing us to decide what in the world to attend to and what to ignore. Similar to spatial data that represent only a selection of geographic information, the ontological commitments we make constrain the domain we wish to model.

Reasoning in machines is a computational process and therefore requires choosing a formalism to represent this information. This is made possible by mapping the set of ontological concepts into a set of language constructs. Constraints such as properties and relations that relate ontological concepts to one another determine the combined use of language elements. An interpretation of these elements (i.e., semantics) assigns to each language construct an ontological interpretation (Evermann and Wand, 2005). The formalism should be appropriate in two respects, it has to faithfully render the available data and make the kind of reasoning needed possible. Great care must be taken to define the concepts and relations on an appropriate level of expressiveness. The terms have to be general enough to allow the annotation of all information sources, but specific enough to make meaningful definitions possible. Spatial reasoning especially requires representation at a high level (e.g. Vieu, 1993).

Although we end up with a specific representation language to implement the model, the essential information is not the form of this language but the content, that is, the set of concepts offered as a way of thinking about the world. However, representation and reasoning are inextricably intertwined so that we cannot talk about one without also

unavoidably discussing the other (Davis *et al.*, 1993). On the one hand, knowledge representation plays the role as a medium for pragmatically efficient computation. This efficiency is supplied by the guidance that a representation provides for organising information to facilitate making the recommended inferences. On the other hand, it is important to capture and represent the richness of the natural world. Either end of this spectrum offers its problems: We can ignore computational considerations at our peril, but we can also be overly concerned with them, producing representations that are fast but inadequate for real use.

This chapter is about formalising the conceptual model discussed in the previous chapter with a knowledge representation language. So far, I have considered the term inference in a generic sense to mean any way to get new expressions from old ones. This chapter is concerned with representations that enable sound logical inferences based on asserted knowledge. The literature suggests many different paradigms for knowledge representation formalisms, ranging from formal logic, fuzzy logic, frames, semantic nets, to production systems (Koch *et al.*, 1997). The family of description logics (DL) provides the foundation for these representation tools by offering rich schema languages. In particular, description logics have become an accepted standard for decidable knowledge representation. They play an increasingly important role for building the next generation of deductive, ontology-based information systems (Wessel and Möller, 2006). The World Wide Web Consortium, for example, endorses the description logic based ontology language OWL (Web Ontology Language) as a standard for ontology representation in the Semantic Web (McGuinness and van Harmelen, 2004). In the next section, we will learn why logic-based languages are useful for the formal representation of knowledge and automated inference. Section 6.2 describes the requirements for translating the conceptual model into a formalised representation language, and why description logics offer the most suitable formalism for the model's implementation.

6.1 Logical foundations for the conceptual model

Logic has its historical origins in Aristotle's efforts to accumulate and catalogue syllogisms in an attempt to determine what should be taken as a convincing argument. A famous rule is the following syllogism: All men are mortal. X is a man. Therefore, X

is mortal. If we assume the truth of the premises, namely the first two sentences, then the law of the syllogism assures us that the third sentence is true whatever the identity of X (Ben-Ari, 1993). For instance, if X is a specific man such as Socrates, we can deduce that Socrates is mortal. This thought continued with Rene Descartes, whose analytic geometry showed that Euclid's work on the logical organisation of geometric principles into axioms and theorems could be married to algebra. By the time of Gottfried Wilhelm von Leibniz in the seventeenth century, the agenda was to seek a calculus of thought – one that would permit the resolution of all human disagreement. In the nineteenth century, Boole provided the basis for propositional calculus with his Boolean algebra (Boole, 1848), which later, with additional work from Frege and Peano, provided the foundation for the modern form of predicate calculus. In the twentieth century, Davis, Putnam, and Robinson took the final steps in sufficiently mechanising deduction to enable the first automated theorem provers (Davis *et al.*, 2003; Zegarelli, 2007).

In the logicist tradition, intelligent reasoning is taken to be a form of calculation, typically, deduction in first-order logic. A formal system is a mathematical model of reasoning based on the syntactic manipulation of sentence-like representations. A sentence has an underlying logical form that represents its meaning. Reasoning involves computations over these logical forms. Logic therefore allows us to model and reason about premises based on the notion of proof. If you want to know whether a particular argument is deductively correct, you can find out by taking its premises as given and then trying to derive its conclusion by applying a specified set of rules. If a proof or a derivation is possible, then the argument is deductively correct, that is, the conclusion is deducible from the premises. In other words, the proof lies in the truth of a sentence and the validity of an argument. Therefore, in contemporary logical theory, the deductive correctness of an argument is a matter of the relationship between the truth of the premises and the truth of the conclusion.

The procedure of checking the statement's validity or satisfiability is called proof-theory, and includes the axioms and rules of inference that state entailment relationships among well-formed formulas (Gašević *et al.*, 2006). Standard model-theoretic semantics assigns truth-values or interpretations to atoms and formulas. An interpretation determines the meaning of a sentence stating that the world is this way

and not that way. Hence, semantic statements can be true or false. A sentence is true under a particular interpretation if the state of affairs it represents is the case. The truth therefore depends both on the interpretation of the sentence and on the actual state of the world (Russell and Norvig, 1995). Finding the truth-value of an arbitrary statement is a matter of examining the truth-values of the variables and the truth tables of the logical operators in component parts of the statements. These logical operators (often called Boolean operators after their inventor George Boole (e.g. Boole, 1848)) are known as truth-functional connectives. They build up the truth-value of a complex sentence by using the operators to connect simpler sentences (Barwise and Etchemendy, 1999). For example, negating a true statement turns it into a false statement. Negating a false statement makes it true, that is, every statement has the opposite truth value from its negation. The symbols $\wedge$, $\sqcap$ are used to express conjunction in our language (usually expressed with terms like *and*, *moreover*, and *but*). It implies that a statement built from this constructor is true only when both parts of the statement are true. Otherwise, it is false. The symbols $\vee$, $\sqcup$ express disjunction in our language, equivalent to using the word *or*. Because the $\vee$-operator is inclusive, its semantics says that an or-statement is false only when both parts of the statement are false. Otherwise, it is true. The symbols $\rightarrow$, $\supset$ express logical consequence and imply that when an if-statement is true and the first part of it is true, the second part must be true as well. The symbols $\leftrightarrow$, $\equiv$ express logical equivalence, which means that one part of a statement using the if-and-only-if-operator cannot be true without the other (Zegarelli, 2007). Consequently, these operators help us to determine the truth-values of whole formulas or sentences by checking the truth-values of their atoms. This means, if the conclusion is semantically entailed by the premises, then the entire argument is valid (Rips, 1994). For instance, if A and B are false and C is true, then the following formula $(A \vee B) \wedge C$ is false. This is because if both parts of an or-statement are false, then the or-statement is false. This makes the and-statement false, because the and-statement is true only if both statements are true, otherwise it is false.

There are different types of logic depending on what they commit to as primitives, that is, the set of concept and role constructors they provide. The two main branches are propositional logic and predicate logic. The most basic language is propositional logic whose ontological commitments (what exists in the world) are only facts that can be

true, false or unknown. It treats propositions as single units. Predicate logic, on the other hand, makes finer distinctions. It analyses propositions into combinations of predicates, and is committed to facts, objects and relations. Temporal logic is a variation that includes time. Probability theory and fuzzy logic address facts and degrees of truths (i.e., degree of belief between 0 and 1). Classical first-order logic, however, is by far the most widely used, studied, and implemented version of logic (Sowa, 2000).

First-order logic (FOL) is made up of variables, constant- and function-symbols that build terms, relational symbols that are applied to terms to build predicates, and predicates and logical constructors that build whole sentences (Hedman, 2006; Srivastava, 2008). Variable symbols (x, y, z, e.g. $\forall x. Male(x) \lor Female(x)$) represent arbitrary elements of an underlying set. For example, male is a variable ranging over males. The language's individual constants (a, b, c, e.g. $Male(John)$) represent a specific element of an underlying set. For example, John is a constant symbol denoting a particular male. The symbols for functions (f, g, h, e.g. $\forall x. Male(father(x))$) have any number of variables. If f is a function of one, two, or n-number of variables, then it is called unary, binary, or n-ary, respectively. Unlike predicate symbols, which are used to state that relations hold among certain objects, function symbols are used to refer to particular objects without using their names, for example father of x. In addition, FOL admits a restricted form of quantification that is realised through so-called quantified role restrictions, which are composed of a quantifier, a role, and a concept expression. There are two types of quantifiers: The universal quantifier ($\forall$), read as 'for all', and the existential quantifier ($\exists$), read as 'there exists'. Quantified role restrictions allow one to express properties of entire collections of objects, such as the relationships existing between the objects in two concepts (Calvanese *et al.*, 2001).

Description logics refer to the logic-based semantics that is given by a translation into first-order predicate logic. Description logic languages form the core of knowledge representation systems, and range from high polynomial complexity to no longer polynomial but highly expressive languages, as well as offering various kinds of inference services (Neumann and Möller, 2008). Appendix C describes the family of description logics as well as some of the language's preliminaries. For example, in DL, let C and D be concept descriptions, A be an atomic concept and R be a role name, then

the set of $\mathcal{ALC}$ concepts is inductively defined as follows: $C, D \rightarrow A \mid \neg C \mid C \sqcap D \mid C \sqcup D \mid \forall R.C \mid \exists R.C$ (Espinosa *et al.*, 2007). The interpretation of $\neg C$ is the set of all individuals in the domain that do not belong to the interpretation of C. The intersection of two concepts ($C \sqcap D$) is interpreted as the set-intersection of all individuals in the domain that belong to both interpretation of C and the interpretation of D. The union of concepts ($C \sqcup D$) means that individuals in the domain are instances of either C or D. The existential restriction ($\exists R.C$) should be paraphrased by "amongst other things". Therefore, when given $\exists hasChild.Male$, it means that at least one child must be male. This is an open world assumption, where we assume there is always more information than is stated. This type of assumption is different from the closed world assumption, as for example found in databases where the information we have is everything. Whereas a database instance represents exactly one interpretation, as defined by the classes and relations in the schema, a DL knowledge base represents many different interpretations with all its models (Baader *et al.*, 2003). Consequently, if a database cannot find some data, it returns a negative. However, a reasoning procedure in DL makes no assumption about the completeness of the information it is given, and therefore treats absence of information simply as lack of knowledge. Alternatively, the universal value restriction closes the interpretation of the domain: $\forall R.C$ requires that all the individuals that are in the relationship R with the concept C being described belong to the concept C. For example, $\forall hasChild.Male$ means that all children must be of type male, that is, there can be no child member that is not male. The quantified role restrictions are denoted by the letter C and thus extend the DL base language $\mathcal{AL}$ to $\mathcal{ALC}$.

In summary, logic enables a precisely formulated subset of language to be expressed in a computable form (Sowa, 2000). Whereas its syntax defines abstract formulas or sentences in the language, the semantics or intended interpretations define the meaning of sentences, that is, the truth of a sentence in a world. The justification of applied mathematics is that the result of a syntactical manipulation (theorems or computations) can be used in the real world by mapping from syntax to semantics (Ben-Ari, 1993). This mapping, however, depends on the expressive power of a representation language, and is directly linked to the resources needed for computing a solution. The definition of reasoning problems therefore addresses both decidability (i.e., if a problem is solvable) and its associated computational complexity. Decidability of an entailment problem

refers to a terminating algorithm to compute entailment. These algorithms typically are based on the well-known tableau method to test the logical validity of complex propositions in a formula (D'Agostino, 1992). A formula is said to be satisfiable if the algorithm will constructively exhibit a model of the formula. If the argument is invalid, the model is undecidable and does not terminate. The computational cost of finding a proof may be enormous.

In formal logic, inference procedures for the reasoning problems derive results that are logically implied by a set of premises. However, only inferences that are permitted are sound inferences. In other words, logic permits only those inferences that are encompassed by logical entailment in which every model for the axiom set is also a model for the conclusion. For example, a fundamental inference rule is modus ponens, illustrated earlier with the syllogism that if all men are mortal and Socrates is a man, then Socrates is mortal. It means that if you know that a statement of the form $P \rightarrow Q$ is true, and you know that the P part is true, then you can conclude that the Q part is also true (Zegarelli, 2007). Additional inference rules enable greater deductive power. However, careless use of logic can of course lead to inexplicable situations or paradoxes, that is, self-contradictory statements. Consider the following syllogism: Some cars rattle. My car is some car. Therefore, my car rattles (Ben-Ari, 1993). In particular, the use of an imprecise notation such as natural language can lead to claims that false statements are true, or to claims that a statement is true, even though its truth does not necessarily follow from the premises. Nevertheless, logic and its notion of inference has a number of important benefits, including being intuitively satisfying (a sound argument never introduces error), explicit (so we know precisely what we are talking about), precise enough that it can be the subject of formal proofs, and old enough that we have accumulated a significant body of experience (Davis *et al.*, 2003).

Logic is especially useful because it formally addresses the relationship between representation and the world (Wilson and Keil, 1999). Representation alone is generally not enough. We want to be able to access and process the represented knowledge. Logic achieves this by reducing reality to a set of abstractions, called a model, by working within this model to reach a conclusion, and then applying this conclusion back to reality again (figure 27). This process is most successful when a good correlation exists between the model and reality and when the model lends itself well to the type of

calculations that logic handles comfortably. Although logic is the only well-developed system for assessing the deductive correctness of arguments, the idea that formal logic bears a close relationship to human reasoning is extremely controversial within cognitive science (Rips, 1994).
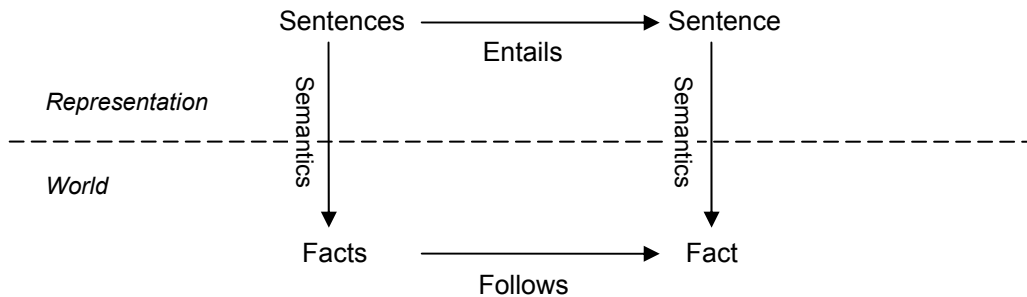
Figure 27 Drawing conclusions from propositions (Russell and Norvig, 1995)

To enable logical reasoning for our conceptual model, we need to specify the terminology of the ontology with first-order logic (Grüninger and Fox, 1995). With its precise mathematical formulation of the properties and relations of entities and proposed axioms about entities in question, a formal language based on logic provides the necessary framework to represent information in an especially useful way and to make it more easily accessible to automated information processing. Yet, simply proposing a set of objects alone, or proposing a set of ground terms in first-order logic does not constitute an ontology. Axioms must be provided to define the semantics, or meaning, of these terms, followed by sanctioned inferences. The commitment to a particular view of the world depends on the choice of a representation technology and accumulates as subsequent choices are made about how to see the world in these terms (Davis *et al.*, 1993). These choices are reflected by the predicates that represent different ontological commitments of all the relevant things that exist in the subject matter or application (Sowa, 2000).

Ontological commitments specify a set of constraints that declare what should necessarily hold in any possible world. An ontology describes concepts (aka classes), properties of concepts (aka attributes or roles), relationships between concepts, and additional constraints (e.g. role restrictions). Ontologies thus play a key part of a broader range of semantics-based technologies and are a sub-area within knowledge representation. There is a wide variety of ontology languages of which some are more

formal than others (Gašević *et al.*, 2006). Ontologies may be simple (having only concepts), frame-based (having only concepts and properties), or logic-based (e.g. OWL). Each representation technology supplies its own view of what is important to attend to. Each suggests, conversely, that anything not easily seen in these terms may be ignored. For example, logic involves a commitment to viewing the world in terms of individual entities and relations between them. Rule-based systems view the world in terms of attributes-object-value triples and the rules of plausible inference that connect them, whereas frames have us thinking in terms of prototypical objects. The selection will have a significant impact on our approach to the task and on our perception of the world being modelled (Davis *et al.*, 1993).

Ontologies are typically expressed by means of diagrams. For example, the entity-relationship conceptual data model and UML (unified modelling language) class diagrams can be considered as ontology languages. These languages have evolved with the Semantic Web. For example, the resource description framework (RDF) is a language used for representing information about resources on the web. RDF describes these resources in terms of properties and property values. Its statements form sets of triples that consist of a subject, a predicate, and an object. Subsequently, the language was extended with RDF Schemas (RDFS) to enable the expression of classes of resources and the properties used with them. RDF and RDF Schemas are recognisable as an ontology language because of their classes (sub- and super-classes) and properties (range and domain of properties). However, RDFS is too weak to describe resources in sufficient detail, and its non-standard semantics with higher order flavour makes it difficult to provide reasoning support. The recognition of these limitations led to the development of new web ontology languages, such as Ontology Inference Layer (OIL), DARPA Agent Markup Language DAML+OIL, and Web Ontology Language (OWL), which began to include logic-based descriptions (Horrocks and Patel-Schneider, 2003).

The investigations of DL language constructors provided a detailed understanding of the semantics and computational properties of, and reasoning techniques for various ontology language designs (Baader *et al.*, 2006). The marriage with logics provided ontology formalisms with the specification of both syntax and semantics necessary for the use of standard inference engines for reasoning over ontologies (Calvanese *et al.*, 2006b; Uschold and Grüninger, 2004). This understanding led to three OWL dialects of

which two provide decidable reasoning problems. OWL FULL is a union of OWL syntax and RDFS, where RDF semantics is extended with relevant semantic conditions and axiomatic triples. Because OWL-FULL provides features that go outside of the description logic paradigm, such as meta-modelling, blending objects and data types, unusual syntactic forms, etc., it does not guarantee computational completeness and decidability. OWL-DL on the other hand is restricted to description logics. It has standard first order model theoretic semantics. This makes it the most expressive of the three sublanguages in that it does not compromise completeness and decidability. Its underlying description logic is $\mathcal{SHOIN}^{(D)}$ (Horrocks and Sattler, 2005). The different letters in the name stand for the sets of constructors available in the language. Hence, the language restricts the form of number restrictions to be unqualified, supports role hierarchies, nominals and inverse roles, and adds a simple form of data types (often called concrete domains in DL). OWL-Lite is an easier to implement subset of OWL-DL with less expressive power being based on $\mathcal{SHIF}^{(D)}$. The W3C Web Ontology Working Group considered the design of simpler ontology languages with more tractable inferences important (Baader *et al.*, 2006). OWL-DL and OWL-Lite are thus by far the most used languages in ontology applications.

OWL exploits a considerable existing body of description logic research (Horrocks, 2005a). Its specific syntactic constructs are written as combinations of RDF syntactic constructs (Horrocks and Patel-Schneider, 2003; Schwitter and Tilbrook, 2006). As a result, OWL relies on XML for syntax and is semantically layered on top of RDF/ RDFS from where its three sublanguages borrow different sets of constructors, which affect their expressive power. It provides a source of sound and complete algorithms and optimised implementation techniques for deciding key inference problems, and therefore is used in implemented DL systems to provide necessary reasoning support. OWL evolved to a standard ontology modelling language, which led to the notion of ontology being treated as a synonym for a description logic knowledge base (Calvanese *et al.*, 2006b). In particular, the standardisation of OWL led to the development and adoption of a wide range of tools and service, including reasoners such as FaCT++, Racer, and Pellet, and editing tools such as Protégé (Baader *et al.*, 2006). Although OWL was initially designed for the Semantic Web, it is now widely used in ontology development in general. This means, its language constructs are being continuously

extended by exploiting the ever increasing developments for more expressive languages.

The advantage of using off-the-shelf software is that it is accessible to everyone and illustrates what solutions are practically achievable. However, by using OWL as the implementation language, we have to accept the language's restrictions in terms of how and where language constructs can be used to guarantee decidability. For example, the treatment of specific domains with fixed (concrete) semantics is challenging for description logics. Under certain conditions, objects and relations of a concrete domain, such as space, can be built into a description logic so that knowledge representation and reasoning can be performed with other than purely symbolic objects. Concrete domain reasoning is still actively explored, including the coupling of geometric computations such as topology with symbolic reasoning services (e.g. Gütter *et al.*, 2008), as well as extending data type expressivity in the next generation of OWL2 (e.g. Motik and Horrocks, 2008; Cuenca Grau *et al.*, 2008). In principle, spatial representations are possible with expressive spatial concrete domains (e.g. Möller *et al.*, 2000). Research in spatial reasoning provides us already with logical calculi for representing and reasoning with qualitative spatial relations over regions (e.g. Cohn *et al.*, 1997; Bennett, 2001; Galton, 1999; Wolter and Zakharyaschev, 2005; Bittner and Stell, 2000; Haarslev and Möller, 1997; Isli *et al.*, 2001; Lutz and Möller, 1997; Möller and Wessel, 1999). In OWL, however, these calculi are not yet implemented, and the concrete domain is restricted to expressing only some quantitative properties. Despite these limitations, description logic based languages offer a suitable formalism for implementing the conceptual model, as we will see in the next section.

6.2 Model-based recognition of the dwelling and beyond

A conceptual model captures expertise in an informal, but structured way. It describes the different types and roles of knowledge in reasoning tasks. A formal model encodes this knowledge in a symbolic formalism with a mathematically sound basis and a declarative semantics. It allows eliminating ambiguities and inconsistencies from the conceptual model and enables formal verification and validation (Benjamins *et al.*, 1999). We now need to transform the content of the conceptual model into a formal model. This translation is essentially a mapping between two languages, or media of

expression, that preserve certain aspects but not others, that is, leaves them invariant (Kuhn, 2004a). The key question is what to preserve and what to lose in the process. We need to determine the appropriate mappings from the source conceptualisation to the target language. This is not a trivial task. Poli (1996), for example, notes that the ontology and the logic (or at least the formalism), which should give it formal rigour, lie at different levels that should not be confused. In passing from the ontological tree to the logical tree, changes may occur of which one should be aware, and there is nothing to guarantee the neutrality of the translation. Furthermore, there may be different logical translations of the same ontological structure, which also may not prove to be compatible with the entire ontology.

So far I have treated 'concept' as a linguistic artefact, where it is used in place of a name or word as a device that allows us to abstract away from incidental syntactic differences and focus instead on those sorts of relations between terms which are important for reasoning. On the engineering reading, concepts are creatures of the computational reality, which exist through their representations in software, or in systems of axioms (Smith, 2004). Concepts are conceived as universals to which the general terms used in making assertions correspond. Universals are repeatable, abstract, and lack specific locations in space-time. For example, the concept Public House is a universal concept. Particulars are the instances of such universals, which exist in the real world of space and time. Manchester House, for example, is a physical instance of the concept Public House. A universal is defined as anything that is instantiated, and an instance as anything that instantiates some universal. The term universal thus signifies what the corresponding instances have in common. The relation of instantiation is hereby taken as primitive, and it is specified axiomatically that it holds exclusively between instances and universals. To support semantic annotations we need to define the necessary and sufficient conditions an information entity (particular) has to fulfil to belong to a concept (universal) (Visser *et al.*, 2000). Indeed, any theory can be formulated in many different ways, which can take different sets of concepts. Some choices may be easier to work with than others, depending on the conceptual vocabulary one wants to formalise within the theory. However, the possibility of defining one concept in terms of others gives a very powerful mechanism for organising and streamlining ontology development.

This two-level representation with concepts, universals or high-level knowledge (i.e., conceptual representation) on the one side, and asserted particulars, instances or individuals (i.e., factual representation) on the other side has been addressed throughout this thesis. Figure 28 illustrates how the two representations relate to one another, and how the representational framework classifies instances into higher-level classes. The figure is an adaptation of Neumann's (2005) model for scene interpretation to that of topographic data. The underlying idea is that all high-level structures can be described in a homogeneous way as composite entities with spatially related parts. The shaded areas in figure 28 emphasize how these entities form specific configurations which in turn link to other configurations. The high-level concepts explicitly define the constituting elements and their characteristics. Reiter and Mackworth (1989) were the first to show that scene interpretation is formally equivalent to logical model construction. Hence, instead of concluding from the evidence that this is, say a residential area, the conceptual model of a residential area explains the evidence with its composition that builds on a declarative representation of knowledge. Since the creation of a configuration requires abstraction, it should provide a set of guiding principles that select, organise and order relevant elements independent of contingent factors (Pesaresi and Bianchin, 2000). As explained in the previous chapter, the conceptual representation defines a land use scene consisting of primitive objects that aggregate into higher, more meaningful entities. Interpretation is defined as an instantiation of a conceptual knowledge base consistent with evidence, that is, with information about the scene. In other words, the conceptual model maps onto the evidence given by the data representation, and is implemented as constraint satisfaction. Henceforth, inference is treated as a search problem of classifying possible interpretations defined by the taxonomical and compositional relations and by incrementally instantiating concepts while maintaining consistency.
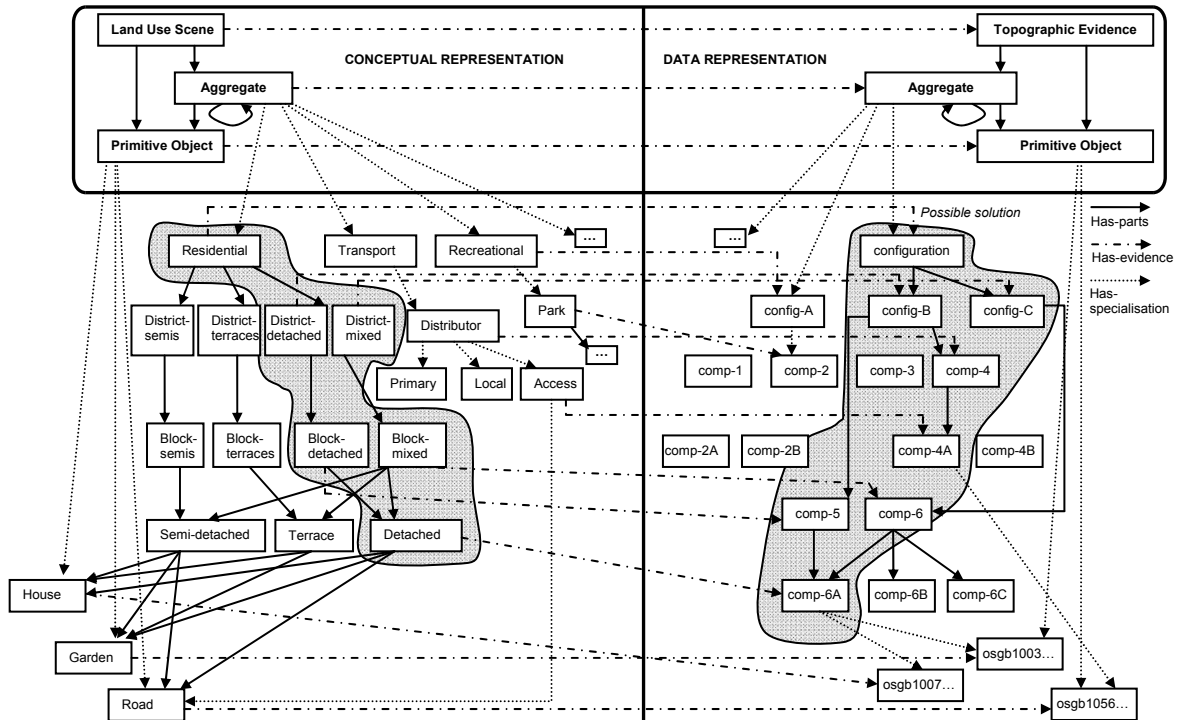
Figure 28 Adaption of Neumann's (2005) model for interpreting topographic data

Knowledge representation technologies provide flexible access to information in many different modalities (Bresciani and Franconi, 1996). They provide the organization, the classification, and the conceptual modelling of information. They aggregate and abstract data in various dimensions and at different levels of granularity. However, the choice of the knowledge representation language rests on the inference mechanisms needed by the application that uses the ontology. We require a representation formalism that not only allows us to describe simple taxonomic relationships, but also provides suitable axioms to express other relationships between concepts and to constrain their intended interpretation. For example, the subsumption or taxonomic inclusion allows us to express that a terraced house is a kind of house. Instantiation means that a topographic feature with the identifier osgb10000040376335 is an instance of terraced house. The individual part-of relation allows us to say that a garden is part of a terraced house, whereas the membership relation states that this house is a member of the collection of houses in the block of terraced houses. The partonomic inclusion between universals offers statements such as every instance of the universal terraced house is an individual part of some instance of the universal residential area. The partition of (or subdivision of) relations expresses that the collection of blocks of terraced houses forms a partition

of districts of terraced houses. Currently, only with description logic based languages, the inference engine (reasoner) can infer these relations at run time.

With the formal, logic-based semantics of description logics, we have the expressiveness for modelling the domain as well as the necessary reasoning services that make automatic inferences over our knowledge base. Reasoning is a central service that allows one to infer implicitly represented knowledge from the knowledge that is explicitly contained in the knowledge base. The capability of exploiting the description of the model to draw conclusions about the problem at hand is a particular advantage of modelling using DL. In addition, a DL system offers the components to store both the necessary types of representations: The conceptual model consisting of a set of terminological axioms and the domain specific asserted facts. These two components are traditionally called the TBox and ABox of a knowledge-based system, as briefly mentioned in chapter five. The TBox equates to an ontology, which contains intentional knowledge in the form of a terminology (taxonomy/partonomy) consisting of atomic concepts (unary predicates) and attributes, usually called roles (binary predicates). These are built through declarations that describe general properties of concepts, resulting in a lattice-like structure entailed by the subsumption relationship. The resulting hierarchy of assertions forms the representational structure for the conceptual model.

The assertions in the TBox are restricted to so-called definitions. A definition is an assertion stating that the extension of a concept denoted by a name is equal to the extension of another complex concept (Calvanese *et al.*, 2001). These statements take the form of terminological axioms expressed as $A \sqsubseteq C$ (primitive concept definition/ concept inclusion) and $A \equiv C$ (concept definition/ concept equation). A primitive concept is an atomic concept occurring only on the right-hand side of axioms. The defined concept is an atomic concept occurring on the left-hand side of an axiom (Baader *et al.*, 2003). TBoxes differ from each other by the type of TBox-statements they allow (Donini *et al.*, 1996). A primitive concept definition in the form of an inclusion assertion $A \sqsubseteq C$ states a necessary but not sufficient condition for membership in the class A. By means of C, the assertion specifies only necessary conditions for an object to be an instance of the atomic concept A. Although the

property C is necessary for an object to be in the class A, this condition alone is not sufficient to conclude that the object is an instance of class A, unless it is explicitly stated. Symmetrically, an assertion $C \sqsubseteq A$ specifies a sufficient condition for an object to be an instance of A. In contrast, an equality assertion $A \equiv C$ states both necessary and sufficient conditions for membership in the class A. It corresponds to the pair of assertions $A \sqsubseteq C$ and $C \sqsubseteq A$. This means that besides having the property C, it is necessary for an object to be in the class A. For example, we can define the concept detached house by stating that a detached house is a house that does not touch some other house, e.g. $DetachedHouse \equiv House \sqcap \neg\exists touches.House$. An individual only then becomes a member of the class DetachedHouse when it meets both necessary and sufficient conditions of being a house and not touching some other house. Equality assertions are typically used to define a taxonomy of concepts. It is assumed that each atomic concept may appear at most once on the left hand side of an assertion to ensure the taxonomy is acyclic. Other forms of expression are $R \sqsubseteq S$, $R \equiv S$ and $R^+ \sqsubseteq S$, where R, S are roles, and R^+ is a set of transitive roles. A set of axioms of the form $R \sqsubseteq S$ where both R and S are atomic is called role hierarchy, and its presence is usually indicated by appending $\mathcal{H}$ to the name of the DL (Baader *et al.*, 2003). Reasoning tasks reason on the concept expressions obtained by unfolding the definitions, whereby replacing atomic concepts on the left hand side of a knowledge base assertion with the corresponding right hand side (Calvanese *et al.*, 2001).

The ABox, on the other hand, contains extensional or assertional knowledge that is specific to the individuals of the domain of discourse (i.e., the evidence in terms of topographic features). The ABox is a set of assertions that is realised by permitting concepts and roles to be used in assertions on individuals. For example, with the concept membership assertion $C(a)$, where C is a concept name and a is an individual, we can express that the topographic feature OSGB1000040381257 is an instance of the concept building: $Building(OSGB1000040381257)$. Further, we can assert that this topographic feature touches another feature by using the role membership assertion $R(a, b)$, where a, b are individual names and R is a role name: $touches(OSGB1000040381257, OSGB1000040381258)$.

What is required are mechanisms that feed concrete data, e.g. topographic features, into the ABox (Neuman and Möller, 2008). For example, a quantitative description of the map scene consists of a list of all primitive objects present, each described by its unique identifier (e.g. TOIDs), other available attributes (e.g. descriptive terms), and calculated spatial relations. A spatial analysis provides measures of the distribution of physical and other spatial structures in the map. We can then apply predicates in qualitative primitives that correspond to notions such as near, far or touching, whereas map elements constitute all constant symbols of an interpretation. In other words, lower-level processes will supply data for instances of concepts, which are modelled as parts of the map scene. Context information may be entered into the ABox in terms of instantiated aggregates, which constrain other possible map objects. For example, if the context of a residential area is given, it is assumed that a corresponding aggregate is instantiated and possible parts, such as terraced or semi-detached housing, are expected as constituents of the interpretation. Such context-based instances help to guide the interpretation process. Hence, the ABox will contain concrete facts about the map data, i.e., its individuals, context information in terms of partially specified concept instances, as well as the resulting high-level description as generated by the inference process.

Protégé allows to implement such a system through its OWL-DL language with which we can specify the concepts of the TBox as classes and the concrete facts of the ABox as individuals. Since a DL system offers standard inference procedures for both TBoxes and ABoxes, we can reason over the defined classes and their individuals and infer implicit, new information. As described in the previous section, logical inference is a process that implements the entailment relation between sentences. A reasoner evaluates the truth of sentences with respect to a model of formally structured worlds. It checks if knowledge is correct and meaningful, that is, if classes have instances. It checks that knowledge is minimally redundant, i.e., that there are no unintended synonyms. No human intervention is required to spot glitches in reasoning. Further, a reasoner can query knowledge. Query answering is performed simply by iterating instance checking for all the individuals in a knowledge base. By means of ABox reasoning and a store of asserted descriptions of individuals, a DL system can query which individuals occurring in the assertions are instances of some concept description (retrieval), or alternatively, given an individual a, what is the most specific concept C in the TBox that a is an instance of (realisation) (Bechhofer et al., 2003; Baader et al., 2003). Answering queries

in DL systems therefore goes beyond query answering in relational databases, because it must consider all models and requires deduction (Esposito *et al.*, 2007). In addition, subsumption ensures that the right place for a concept C is found in the taxonomy implicitly present in a TBox. It verifies whether a generic subsumption relationship between concept expressions is a logical consequence of the declarations in the TBox, thus ensuring consistency of the ontology. The task of inserting new concepts in a taxonomy is referred to as classification. Here, the reasoner determines for a given concept C in a TBox whether the new concept D subsumes C, or D is subsumed by C. Table 7 summarises the reasoning tasks with respect to the ABox and TBox (Calvanese *et al.*, 2001; Baader and Küsters, 2006).

Table 7 Reasoning tasks for the TBox and ABox

TBox	ABox
Inferencing of relationships, be they transitive, symmetric, functional or inverse to another property. **Equivalence of concepts** within a terminology is the problem of deciding whether two concepts are logically equivalent ($C \equiv D$). **Subsumption** checks whether one concept is more general than another. **Satisfiability** generally is the problem of checking whether a knowledge base has a model, i.e., a valid interpretation. Concept satisfiability is the problem of checking whether concept C is satisfiable with respect to a knowledge base. **Classification**, which places a new concept in the proper place in a taxonomic hierarchy of concepts. **Concept consistency** is the problem of deciding whether concept C is consistent in a knowledge base. **Logical implication** is the problem of deciding whether a knowledge base implies an inclusion assertion $C \sqsubseteq D$, which is whether a generic relationship is a logical consequence of the declarations in the TBox.	**Consistency checking** of instances. **Entailments**, which are whether other propositions are implied by the stated condition. **Satisfiability** checks that the conditions of instance membership are met. **Infer property assertions** implicit through the transitive property. **Instance checking** is the problem of checking whether the concept membership assertion $C(a)$ is satisfied in every model of a knowledge base. It verifies whether a given individual is an instance of a specified concept. **Knowledge base consistency**, which is to verify whether all concepts admit at least one individual. In other words, to check whether a given ABox is consistent with respect to a TBox **Realisation** is the problem of checking the most specific concept C in the TBox that an individual a is an instance of. **Retrieval** is the problem of checking whether an individual is an instance of some concept description C.

There are well-established and optimised reasoning algorithms for these kinds of reasoning tasks, such as the earlier mentioned tableaux procedures (e.g. Baader and Sattler, 2001; Möller and Haarslev, 2008). The tableau calculus is specifically designed for solving the problem of satisfiability, validity and entailment by incrementally

building a model of a given formula, decomposing it in a top/down fashion, and exhaustively looking at all possibilities until it can eventually prove whether a model exists or not. In other words, tableau rules correspond to constructors in logic and stop when no more rules are applicable or a clash occurs. If a clash occurs then the problem is said to be undecidable. However, the property of (un)decidability lies with the general property of the problem and not of a particular algorithm solving it. An algorithm is just a computational process that uses a finite number of steps to solve a particular problem. The proof procedure's time for solving a problem depends on the complexity of the formulas. Although decidability of a formal language can be achieved by restricting it (Calvanese *et al.*, 2001), we have to sacrifice some of the logic's expressive power and with that the complexity with which we can describe a problem. However, given a decidable problem, the issue of computational difficulty in solving the problem remains in terms of the use of computing power and resources. For example, validity testing for sentential logic is equivalent in computational complexity to problems for which every known algorithm requires an amount of time equal to some exponential function of the length of the problem statement (Rips, 1994). Therefore, the complexity class to which a problem belongs is again a general property of the problem and not of a particular algorithm solving it. In general, the more restricted the representational power, the faster is the inference.

The use of logic and its inference scheme have been much researched in computer vision (e.g. Zhang, 2007). Despite limitations in expressivity and computational power, formal logics provide us with the necessary means of modelling and reasoning about geographic space in an explicit and more natural way. Indeed, there are different ways to represent knowledge. Not only is there a large variety of languages to choose from, but there are also other systems such as Minsky's frame theory (Minsky, 1975), probabilistic models (e.g. Bayesian network), or associative structures and cognitive learning paradigms (e.g. neural network). However, modern ontology languages based on description logics offer well-proven, standardised representation and reasoning mechanisms. The question that remains is how the language's expressivity and reasoning ability will live up to the requirements for deriving a high-level representation of land use from topographic data.

Conclusions

Modelling geographic space in terms of its context and arrangement is a difficult task. Its spatial patterns were never consciously created in the first place, thus making it a challenge to consciously recreate these patterns through a system of generic design guidance (Marshall, 2005). Although a conceptualisation of space provides the necessary abstraction, it requires guiding principles to which relevant elements are to be selected, organised and ordered. Ontologies provide the necessary means to structure the ontological commitments of our domain, and their semantic-based languages provide the necessary inference services to make new knowledge explicit. This chapter showed how the conceptual model can be made accessible to the computer through knowledge representation formalisms. The model's rich descriptions of meanings are explicitly expressed by an ontology. The ontology not only provides the vocabulary of terms, through which new terms can be formed by combining existing ones, but it also formally specifies the semantics of a shared domain. Due to its logic-based semantics a machine can reason about the asserted knowledge and infer higher level, initially implicit information. Since interpretation is a cognitive as well as knowledge-intensive task, a knowledge-based approach seems to lend itself to the problem of inferring additional information from topographic data. An important lesson to emerge from controversies around knowledge representation is that the representation of knowledge cannot be completely isolated from its hypothesized functions in cognition (Wilson and Keil, 1999). Just as knowledge representation paradigms have proven useful for computer vision (e.g. Möller *et al.*, 1999), so should they be useful for the interpretation of geographic information.

Knowledge representation formalisms interface to common-sense knowledge and represent conceptual models with well-defined semantics that exploit validated inference procedures. Description logics in particular constitute a whole family of formalisms that have obtained much attention in the last decade. A description logic's object-oriented knowledge representation is similar to frame systems used in many knowledge-based application systems, but it is based on formal semantics. Because description logics realise a subset of first-order logic, they guarantee the decidability and correctness of reasoning services including consistency checking, subsumption, satisfiability, classification, abstraction and instance checking and retrieval (Neumann

and Möller, 2008). Although logic offers complex properties and highly optimised implementations, such as OWL, it is important that all the available knowledge can be fully axiomatised and represented using such formal languages. It is also required that for the used formal language there exists an elaborated model theory, and that for the purpose of reasoning the language is decidable. This leads to the important consideration of the trade-off between expressiveness of a terminology and the complexity of reasoning services, which ultimately has an effect on the representation of the problem. Unfortunately, desirable features may easily lead to undecidability. For example, concrete domains must be incorporated to support spatial reasoning, which is not yet fully available in OWL.

The trade-off between expressive power and deductive complexity is a central issue of knowledge representation formalisms. DLs have been thoroughly investigated especially with the aim of devising knowledge representation formalisms with a good compromise between expressive power and complexity of reasoning (Calvanese *et al.*, 2001). Even though classical first order logic has enough expressive power to define all of mathematics and the semantics of every version of logic, including itself, logic has its own limitations. It may be able to define fuzzy logic, modal logic, neural networks, and even higher-order logic. It may be the best-defined, least problematic model theory and proof theory, and it can be defined in terms of a bare minimum of primitives: Just one quantifier (either $\forall$ or $\exists$) and one or two Boolean operator (Sowa, 2000). However, all sentences in logics are assertions, and reasoning based on formal logics is limited to deriving truth-values and proofs for such assertions. Hence, it is difficult to model human reasoning that involves assumption, likelihood, belief, doubt, etc.

Logical computation involves regimenting arguments in ways that are often unintuitive. Therefore, it remains questionable how relevant logical proof theories are to the study of human reasoning (Rips, 1994). Formalisms such as semantic or associative networks in AI as well as the classical relational calculus or first-order predicate calculus cannot represent or accommodate inexactness. Even with fuzzy logic, the calculation of a quantified, statistical probability has by itself a distinct air of artificiality (Peuquet, 1988). This means, when it comes to representing our problem, which stands in some sort of isomorphism to corresponding entities in reality, we are faced with additional

trade-offs between method of structural information handling, the nature of the raw data, the characteristics of the landscape and robustness and processing time of the method (Pesaresi and Bianchin, 2000). Nevertheless, the special-purpose heuristics of this approach will take advantage of particular types of rules or lines in the proof.

On the positive side, first-order logic provides a powerful representation and reasoning system upon which many knowledge representation formalisms are based. The separation between syntax and semantics is one of the major strengths of modern logic (Fagin *et al.*, 2003). They are formally well founded and are suitable for machine implementation. Logic is well understood as it has been extensively studied (it goes back thousands of years to philosophers such as Aristotle). It continues to be actively researched in terms of extending the expressivity of concept languages, the decidability and tractability of inference services, and the integration of predicates over concrete domains. There exist several commercial and experimental developments of DL systems, among them KL-ONE, CLASSIC, LOOM, Racer, and Protégé (e.g. Calvanese *et al.*, 2007a; Duineveld *et al.*, 1999), which can be readily used for implementing ontologies and knowledge bases.

In particular, description logics offer a useful paradigm for modelling our problem, since we are interested in the symbolic processing of high-level interpretation and vision tasks (Möller *et al.*, 1999). Whereas the TBox of a DL system contains sentences describing concept hierarchies, that is, the relations between concepts, the representation of factual knowledge is achieved through the declaration of knowledge about individual objects in the assertional knowledge base (ABox). With the ABox, it is possible to express conceptual properties of instances and relations between individuals, for example, of the contents of a particular map scene. The TBox background knowledge determines what can be inferred from the explicit declarations in an ABox. Using the ABox reasoning services, an ABox individual can be shown to be an instance of certain TBox concepts (instance checking), as well as the set of most specific concept names of which an individual is an instance can be computed (object classification). These services have benefits over traditional database query languages (such as SQL) in so far that the modelling comes more natural (it is easier to construct queries), and new information can be inferred from a given set of information. Therefore, we can capture implicit information that does not exist on the level of pure geographical features by

using background information that is usually well known by humans (Heinzle and Sester, 2004).

Another advantage is that DL provides standardised reasoning services. Instead of programming an object recognition procedure, we can use the object classifier of a DL. This provides significant economical advantages in terms of reusable software components that can be used instead of complex application-dependent software (Möller *et al.*, 1999). The success of applied description logics for pattern classification tasks has been shown previously (e.g. Liedtke *et al.*, 1997; Mayer, 1999). The idea of conceptually defining classes in terms of sufficient conditions, which must be fulfilled by image features, has been successfully applied in computer vision as well as the classification of remotely sensed imagery. This thesis also applies the knowledge representation framework for model-construction in the logical sense by treating the problem of modelling land use as a configuration task. The importance lies in the interface between a GIS and a knowledge base to facilitate both the necessary high-level background knowledge as well as situational context given by the data. This leads to a duality of the generality and application of the problem. On the one hand, the independence of symbolic logic formalisms is an advantage with respect to validity and reusability. On the other hand, this independence poses a severe impediment when domain-specific properties and laws, such as dealing with space and time, must be exploited for a task. This especially addresses the incorporation of concrete domains in DL formalisms to accommodate reasoning other than with purely symbolic objects. The current representation of concrete domains and the lack of spatio-terminological reasoning will limit the implementation of our proposed conceptual framework, which builds upon spatial knowledge. However, first-order logic will be around for a long time, and current obstacles are likely to be solved in the future (Russell and Norvig, 1995). Plus, alternative solutions (e.g., frame-based and rule-based languages) have proven not to be perfect either (Gašević *et al.*, 2006).

Chapter 7

Applied Evaluation: Inference of Residential Area

"Implicit information does not exist on the level of pure geographical features, but on the level of the relationships between the features, their extent, density, frequency, neighbourhood, uniqueness and more. This knowledge often is well known by humans with their background information, however, it has to be made explicit for the computer."

–Heinzle and Sester (2004, p.335)

The use of GIS to answer geographical questions will often search for information not explicitly represented in available databases. The challenge of deriving implicit facts from explicit geographic knowledge is mainly a result of the lack of semantics contained in spatial databases (Verastegui *et al.*, 2006). The use of formal ontologies to model, classify and annotate data of various domains has been explored for this purpose (e.g. Villanueva-Rosales *et al.*, 2007; Wolstencroft *et al.*, 2006; Stevens *et al.*, 2007; Bada *et al.*, 2004). As we learnt over the course of this thesis, ontologies are not only helpful as a specification for a required domain, but they provide logic-based search for better information (Uschold and Grüninger, 2004).

To use ontology in the engineering sense, we have to think globally but act locally. We need to think of what it is we want to extract from topographic data in a global sense to get a complete understanding of the domain and provide context for the inference. However, the actual inference takes place locally on a specific set of data, depicting a specific location in reality, which will have constraints affected by local surroundings as in spatial layout. The fact that environments vary in terms of their physical, climatic, and cultural context means that the ontological commitments made locally, do not necessarily apply globally. For instance, the spatial layout of residential areas potentially varies from country to country (chapter 3). Acknowledging these circumstances, the formalised example begins at 'home' with the typical spatial layouts found in Great Britain that are portrayed in Ordnance Survey's topographic data.

The ontological commitments are given by the conceptual model that imposes constraints on the domain. In representing knowledge about the real world, one part of the system is the body of knowledge to be represented. This results in a partial representation of characteristics within the world where the complexity of the problem is reduced both spatially as well as in terms of content (chapter 5). Another part is the representing formal structure (chapter 6), and a third part establishes the relations between the body of knowledge and the formal structure. This chapter establishes this relation by applying existing ontology technology to infer higher-level functional information from topographic knowledge.

To begin with, we make choices about the vocabulary of terms (predicates, functions, and constants) of a domain. The resulting vocabulary, or informal list of the concepts in the domain, is what is known as the ontology of the domain. By writing logical sentences or axioms about the terms in the ontology, we accomplish two things: First, we make the terms more precise so that humans will agree on their interpretation, and second, we make it possible to run inference procedures to automatically derive consequences from the knowledge base. We then encode a description of the specific problem instance, which involves writing simple atomic sentences about instances of concepts that are already part of the ontology. Lastly, we post queries to the inference procedure and get answers, that is, we can let the inference procedure operate on the axioms and problem-specific facts to derive the facts we are interested in knowing (Russell and Norvig, 1995). To understand this process better, we now turn to the implementation by applying the proposed conceptual model in Protégé.

In the next section, after introducing the study areas, the conceptual model is implemented in OWL-DL by asserting necessary knowledge in the ABox and TBox of the knowledge base. This is achieved by defining the concepts of the conceptual model as classes in the OWL ontology, and converting the factual knowledge stored in the database into OWL individuals. Section 7.2 applies the asserted knowledge for concept-based instance retrieval and classification. In particular, the thesis illustrates how a description logic's reasoning services aid the inference from type of dwelling, type of urban block, and type of district to residential area. Section 7.3 discusses the results and evaluates the benefits and current limitations of this approach and its methods, followed by conclusions.

7.1 Implementing the model in OWL-DL with Protégé 4 Alpha

Protégé is currently the leading ontology development editor and environment. The platform was developed at Stanford University and has already been through a number of versions and modifications (Gašević *et al.*, 2006). It facilitates the defining of concepts (classes), properties, taxonomies, various restrictions, and class instances. Its uniform graphical user interface has a tab for the collection of knowledge into a knowledge base conforming to the ontology. Protégé supports several ontology representation languages, including OWL and RDFS. To accommodate the formal logics in OWL, Protégé implements reasoners such as FaCT++ and Pellet that provide automatic inference services including satisfiability, subsumption checking, and instance retrieval (Sirin *et al.*, 2007).

In chapter six, I first introduced OWL. This chapter implements the proposed conceptual model in OWL-DL. I am using OWL because it is freely available through Protégé, and it is the current standard endorsed by the World Wide Web Consortium. As outlined earlier, OWL is an ontology language that provides the formal foundations and reasoning support based on well-defined model theoretic semantics. Its basic constructs are classes (denoting sets of instances), properties (denoting relationships between individuals) and individuals (denoting objects in the world). These constructs are equivalent to the concepts, roles and individuals in first-order logic (FOL).

Figure 29 and figure 30 outline OWL's class constructors and axioms that support the modelling of a given domain (Horrocks, 2006). As explained in chapter six, OWL constructors allow one to specify the intersection of classes by combining two or more classes with the and-operator. In addition, it allows the union of classes with the or-operator, complement classes by negating another class, and restrictions by determining the type and possible number of relationships a class of individuals can participate in (e.g. quantifier, cardinality and has value restrictions). With properties aka roles, we can determine the relationships between individuals. The main categories of properties are object and data type properties. The former links individuals to individuals. The latter links individuals to data type values such as integers, floats and strings. The data type property models the so-called concrete domain. As shown in figure 30, properties can take different characteristics. A functional property can only take one value. An inverse

property denotes the inverse of a relationship (e.g. partOf = hasPart). Inverse functional refers to the inverse of the property that is functional. Symmetric means that if a property links individual a to b then it can be inferred that it links b to a. A property is transitive if it links a to b and b to c, then it also link a to c. These constructors and axioms have been restricted so that reasoning in OWL-DL is decidable.

Constructor	DL Syntax	Example	FOL Syntax
intersectionOf	$C_1 \sqcap \ldots \sqcap C_n$	Human $\sqcap$ Male	$C_1(x) \wedge \ldots \wedge C_n(x)$
unionOf	$C_1 \sqcup \ldots \sqcup C_n$	Doctor $\sqcup$ Lawyer	$C_1(x) \vee \ldots \vee C_n(x)$
complementOf	$\neg C$	$\neg$Male	$\neg C(x)$
oneOf	$\{x_1\} \sqcup \ldots \sqcup \{x_n\}$	$\{$john$\} \sqcup \{$mary$\}$	$x = x_1 \vee \ldots \vee x = x_n$
allValuesFrom	$\forall P.C$	$\forall$hasChild.Doctor	$\forall y.P(x,y) \rightarrow C(y)$
someValuesFrom	$\exists P.C$	$\exists$hasChild.Lawyer	$\exists y.P(x,y) \wedge C(y)$
maxCardinality	$\leqslant nP$	$\leqslant$1hasChild	$\exists^{\leqslant n} y.P(x,y)$
minCardinality	$\geqslant nP$	$\geqslant$2hasChild	$\exists^{\geqslant n} y.P(x,y)$

Figure 29 OWL class constructors

Axiom	DL Syntax	Example
subClassOf	$C_1 \sqsubseteq C_2$	Human $\sqsubseteq$ Animal $\sqcap$ Biped
equivalentClass	$C_1 \equiv C_2$	Man $\equiv$ Human $\sqcap$ Male
disjointWith	$C_1 \sqsubseteq \neg C_2$	Male $\sqsubseteq \neg$Female
sameIndividualAs	$\{x_1\} \equiv \{x_2\}$	$\{$President_Bush$\} \equiv \{$G_W_Bush$\}$
differentFrom	$\{x_1\} \sqsubseteq \neg\{x_2\}$	$\{$john$\} \sqsubseteq \neg\{$peter$\}$
subPropertyOf	$P_1 \sqsubseteq P_2$	hasDaughter $\sqsubseteq$ hasChild
equivalentProperty	$P_1 \equiv P_2$	cost $\equiv$ price
inverseOf	$P_1 \equiv P_2^-$	hasChild $\equiv$ hasParent$^-$
transitiveProperty	$P^+ \sqsubseteq P$	ancestor$^+ \sqsubseteq$ ancestor
functionalProperty	$\top \sqsubseteq \leqslant 1P$	$\top \sqsubseteq \leqslant$1hasMother
inverseFunctionalProperty	$\top \sqsubseteq \leqslant 1P^-$	$\top \sqsubseteq \leqslant$1hasSSN$^-$

Figure 30 OWL Axioms

Given the individuals, classes and properties of an ontology, these are mapped via the interpretation function $\mathcal{I}$ to the interpretation domain $\Delta^{\mathcal{I}}$ based on the semantics of standard first order model theory. Figure 31 shows how the property 'touches' is interpreted as a set of pairs of individuals from the domain Building, and how the class 'SemiDetachedHouse' is interpreted as a set of individuals that is equivalent to sharing a subset of the domain defined as House where an individual touches maximal one House. This means for an individual to be classified as SemiDetachedHouse, it must be in a touch relation to another individual, and both these individuals must satisfy the constraints of the subset of the domain denoted as House.
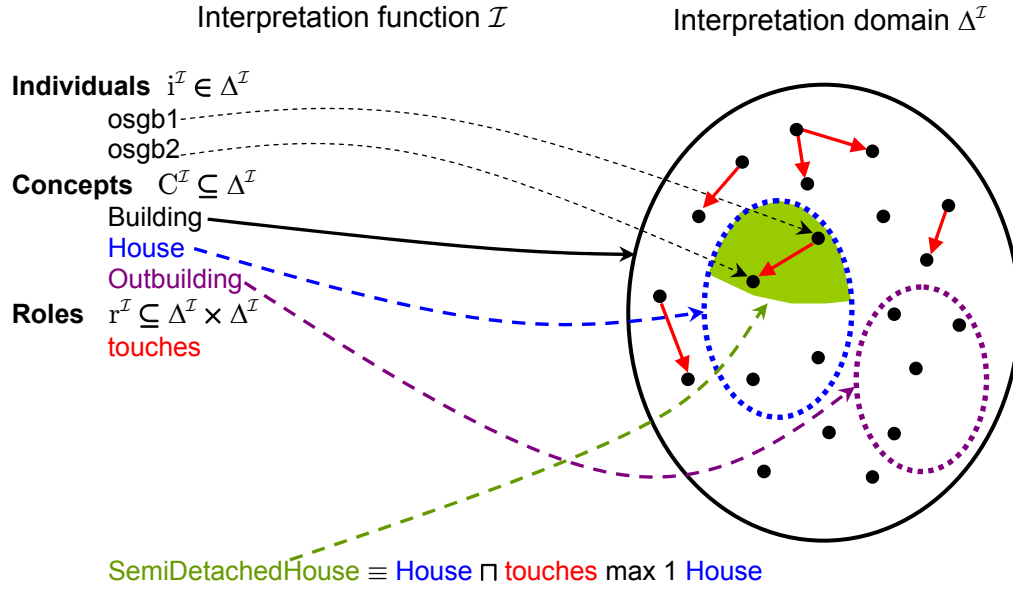
Figure 31 Semantics given by standard FO model theory

The subsequent sections describe the study areas and how Protégé models both database instances as well as the high-level concepts of our conceptual model. In the concrete examples that are going to follow, I am using the widely accepted conventions for writing OWL syntax: Concept names start with an uppercase letter followed by lowercase letters (e.g. SemiDetachedHouse, Building), role names, i.e., properties that relate concepts and individuals, start with a lowercase letter (e.g. hasArea, touches), and individual names are all uppercase (e.g. OSGB1000004037856, URBANBLOCK1). Once an individual is asserted as a member of a specific class, we speak of instances. The Protégé abstract syntax is written in sans-serif typeface.

Study areas

The chosen study area is Glasgow, the largest of Scotland's cities. It is located right on the banks of the River Clyde, situated in the Central Belt of Scotland on the west coast. Glasgow is Scotland's principal commercial centre, and one of the United Kingdom's main regional retail and office centres. Glasgow has a great diversity both geographically and functionally, which makes it a useful study area. Its residential areas are characterised by Victorian architecture, streets of red stone terraced houses with large windows, as well as modern semi-detached and detached properties in and around the city. The three areas Giffnock, Drumchapel and Pollokshields were chosen for the implementation of the conceptual model – each one characterised by different residential properties.

House types and sizes vary in the U.K. An interesting piece of evidence for this is given by RICS Building Cost Information Service's (BCIS) guide to house rebuilding costs (BCIS, 2004). The BCIS guide provides regional rebuilding cost tables broken down by age band, type of house, quality, and size. Table 8 shows BCIS rebuilding cost table for Scotland. BCIS (2004) considers nine Government Office Regions because there are considerable local cost differences within a geographically defined region – Scotland being one of them. It focuses on five major types of house: two storey detached, semi-detached and terraced, and detached and semi-detached bungalows. It represents four age bands, pre-1920, 1920-45, 1946-79, and 1980 to date. These age bands are intended to represent the specification and design typical of the period. Of particular interest are the represented size categories (small, medium and large) for each type of building. The calculations for these sizes have been based on exact areas, which are included in the appropriate rebuilding cost table (table 8). The rebuild cost figures are £/m^2 of gross internal floor area including demolition and fees. The gross internal floor area is the area of the building measured to the internal face of the perimeter walls at each floor level.

Table 8 Scotland rebuilding cost table (BCIS, 2004)

Age	Size	Quality	Detached	Semi-Detached	Terraced	Bungalow	Semi-Detached Bungalow
1980-to date	**Small**	Basic	767	816	777	798	879
		Good	887	946	894	922	1014
		Excellent	1085	1135	1071	1104	1200
	Gross Floor Area m^2		**75**	**52**	**52**	**59**	**41**
	Medium	Basic	646	737	771	731	840
		Good	781	846	877	886	971
		Excellent	987	1000	1031	1146	1159
	Gross Floor Area m^2		**114**	**75**	**64**	**79**	**46**
	Large	Basic	597	705	715	652	713
		Good	762	813	845	852	832
		Excellent	954	1000	1009	1066	1051
	Gross Floor Area m^2		**201**	**130**	**75**	**160**	**79**
1946-1979	**Small**	Basic	793	764	784	774	773
		Good	905	867	895	903	881
		Excellent	1086	1044	1089	1123	1072
	Gross Floor Area m^2		**84**	**83**	**74**	**84**	**69**
	Medium	Basic	734	698	692	713	709
		Good	862	797	790	853	820
		Excellent	1034	983	968	1058	1029
	Gross Floor Area m^2		**110**	**109**	**108**	**111**	**99**
	Large	Basic	615	645	632	636	671
		Good	765	744	723	822	801
		Excellent	939	931	891	1033	1000
	Gross Floor Area m^2		**213**	**135**	**137**	**209**	**114**

Table 8 continued

Age	Size	Quality	Detached	Semi-Detached	Terraced	Bungalow	Semi-Detached Bungalow
1920-1945	**Small**	Basic	879	854	850	820	785
		Good	1050	1025	1021	968	942
		Excellent	1261	1215	1203	1174	1135
	Gross Floor Area m^2		**84**	**72**	**69**	**84**	**69**
	Medium	Basic	841	843	857	770	826
		Good	1012	1008	1023	919	973
		Excellent	1235	1205	1223	1160	1177
	Gross Floor Area m^2		**110**	**91**	**84**	**117**	**99**
	Large	Basic	790	926	948	839	893
		Good	925	1028	1051	977	1005
		Excellent	1125	1238	1258	1228	1211
	Gross Floor Area m^2		**213**	**110**	**106**	**139**	**114**
Pre 1920	**Small**	Basic	928	871	871		
		Good	1100	1039	1045		
		Excellent	1305	1233	1241		
	Gross Floor Area m^2		**104**	**94**	**84**		
	Medium	Basic	903	837	869		
		Good	1058	982	1025		
		Excellent	1257	1174	1224		
	Gross Floor Area m^2		**137**	**137**	**111**		
	Large	Basic	790	825	880		
		Good	959	962	1030		
		Excellent	1152	1175	1242		
	Gross Floor Area m^2		**289**	**189**	**137**		

Size is an important criteria in the interpretation of residential dwellings, as shown in earlier chapters,. Table 8 illustrates the differing sizes between types of houses and their ages. The largest houses are detached houses, especially the older ones. With the increased demand for housing nowadays, the house sizes of modern houses seems to have become smaller over time. Excluding the bungalow type, terraced houses have the smallest gross floor area among semi-detached and detached houses. These size differentiations will proof important for the classifications later on.

Giffnock

Giffnock is an area within East Renfrewshire, Scotland. Its location within Greater Glasgow effectively makes it a suburb of the city. Figure 32 provides an aerial view of the area taken from Google Earth. Giffnock is largely residential in character, surrounded by green spaces. It is a relatively wealthy area, as the Google Street View screen shots confirm in figures 33 and 34. The area consists predominantly of modern as well as Victorian housing. Figure 33 shows some modern semi-detached houses, whereas figure 34 highlights the typical red stone terraces of Glasgow.

Figure 32 Aerial view of Giffnock

Figure 33 Modern semi-detached houses in Giffnock
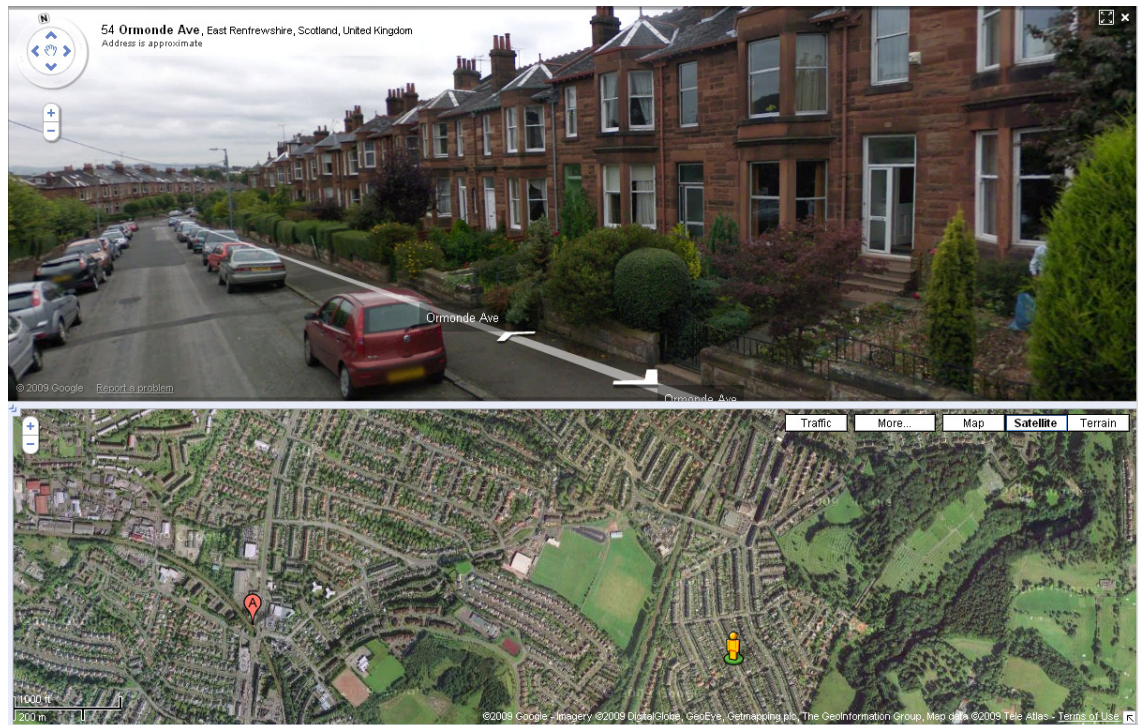
Figure 34 Victorian terraced houses in Giffnock

Drumchapel

Drumchapel is located along the perimeter of the city of Glasgow, bordered by Knightwood and Yoker. As part of an overspill policy, a large housing estate was built there in the 1950s. This post-war social housing scheme suffers from social problems, notably anti-social behaviour and degeneration of its post-war housing. Figure 35 shows an aerial image of the neighbourhood. The area is predominantly of residential character with tall, high-rise buildings, presumably blocks of flats, visible in the centre of the image. Drumchapel is also surrounded by some larger industrial complexes. The Google Street View screen shots in figures 36 and 37 illustrate the post-war and modern houses, mostly flats, terraced and semi-detached houses.

Figure 35 Aerial view of Drumchapel

Figure 36 Modern flats and semi-detached houses in Drumchapel

Figure 37 Post war terraced houses in Drumchapel

Pollokshields

Pollokshields was the first garden suburb to be built in the United Kingdom back in the 19th century. It is among the plushest areas in the city with many avenues of grand Victorian villas accommodating the wealthy. Overall, Pollokshields is an attractive residential area on the south side of Glasgow, just two miles from the city centre. It is a conservation area characterised by substantial sandstone villas and tenements along broad streets. The aerial image in figure 38 shows the spacious layout of the area with plenty of green spaces. Figure 39 gives a Google Street View screen shot of one of the large Victorian villas, and figure 40 shows its typical tenements - large Victorian town houses. These types of houses put Pollokshields in stark contrast to Giffnock and Drumchapel.

Figure 38 Aerial view of Pollokshields
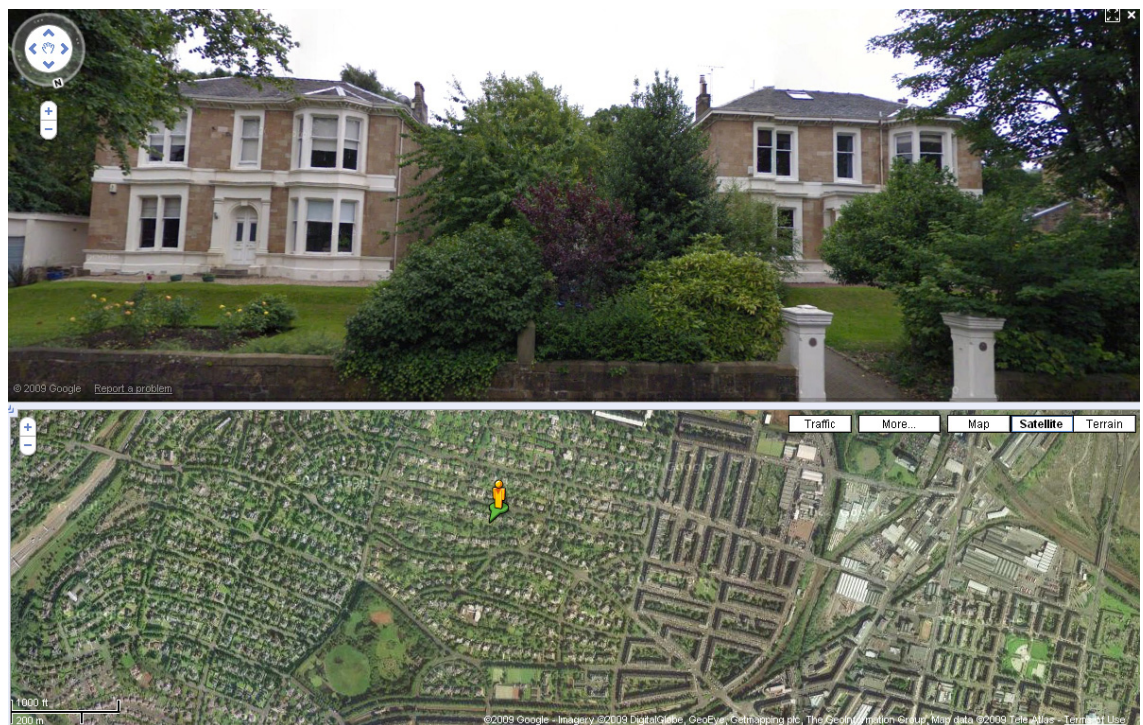
Figure 39 Victorian villas in Pollokshields

Figure 40 Victorian town houses in Pollokshields

Asserting knowledge in the knowledge base

Geographic knowledge in particular is a challenging domain, which requires us to simplify the model by making several restrictive assumptions. When dealing with geographic knowledge, we are dealing with incomplete knowledge. Although this equates to the open world assumption in predicate logic, to effectively deal with the instances of our domain we have to assume complete knowledge, that is, a closed world representation. As elaborated in chapter 6, with the open world assumption the reasoner makes no assumptions about the completeness of the information it is given. However, we need to assume that the information we have is everything so that the reasoner returns an answer. In other words, we need to specify what exists in the topographic map, and we need to assume that the sum of objects in the map are known and finite (Schröder, 1999). This can be achieved by using the universal value restriction, which closes the interpretation of the domain. Furthermore, knowledge is naturally uncertain, especially if obtained from concrete domains such as geography. Even though there is work on incorporating probability theories and fuzzy logics (e.g. Freksa, 1994), for the purpose of the implementation we have to assume that we are dealing with certain knowledge. This is because formalisms only operate on a symbolic level where facts and rules can be postulated to be either true or false. With these restrictions in mind, we

have to find a way to implement the conceptual model in OWL by considering its fundamental notion of incrementally classifying topographic features into higher-level objects from types of housing, blocks and districts to residential areas.

We have two types of knowledge available, that what is contained within the database and that what the conceptual model represents within our domain. This includes:

- General knowledge about objects in the topographic scene as well as the land use domain;

- Implicit contextual knowledge, for example the expert knows that a building in a row of buildings adjacent to gardens is a terraced house;

- Spatial relations between polygons (topology);

- And polygon attributes, such as descriptive group, cartographic text and symbols.

To reason about this knowledge, all information about the map scene must be given a priori in symbolic form, that is, all given knowledge must be axiomatised with logical formulas. This includes the general domain knowledge, which we wish to make explicit within the data, as well as knowledge about the data's topographic features. Respectively, this relates on the one hand to the concepts or classes, whether primitive or defined, that constitute the general knowledge about the domain of interest (TBox). On the other hand, the ABox specifies the topographic data by asserting knowledge about the individuals, which characterise a specific world or situation under consideration, that is, the geographic extent which is being represented and reasoned about. Although the conceptual model guides the types of knowledge required for modelling the domain, we now need to determine precisely the classes and properties in the domain to build our OWL ontology. This means we have to determine domains and ranges for properties, define classes and cardinality restrictions for the relevant properties for each class, and add individuals and relationships as necessary. For the remainder of this section, I discuss how to convert both the general and factual knowledge to OWL.

The TBox

The TBox is the intentional component of a DL system and contains the terminology, or ontology. Here, we create the high-level classes and sub-classes that describe the

information we want to make explicit within the database. Accordingly, we specify classes of our domain of discourse such as House, Building, TerracedHouse, DetachedHouse, BlockDetachedHouses, and so forth. We characterise their relationships to other classes within the taxonomy, thus linking classes to other classes. For instance, through subsumption we specify subclassing mechanisms, such as EndTerracedHouse is-a TerracedHouse is-a Building (figure 41). Other kinds of relationships can be modelled by introducing new object and data properties or roles, such as hasArea, contains, connectedTo, etc. These properties are important to describe both the classes and their individuals. They define the necessary and sufficient conditions for individuals to be classified as instances of certain concepts.

Figure 41 Taxonomy of primitive classes describing the conceptual model

Because concepts are regarded as classes of individuals, we begin with all individuals that are a type of the class Building. This information is explicitly contained within the database; we know which polygons are buildings. For the proof of concept, we will only consider buildings to simplify the reasoning procedure. The aim is to define the higher-level classes to classify the individual buildings. Figure 41 shows the resulting taxonomy of classes according to the proposed conceptual model in chapter five. All classes subsume the abstract class Thing, which is the highest class in OWL. It denotes that everything has an existence in the world.

To define each of the classes, we use Protégé's abstract syntax, which is based on Manchester OWL Syntax (Horridge *et al.*, 2006). The implemented OWL code can be found in appendix D. The definitions have to be reduced and kept as simple as possible to avoid overly complex axioms that slow down the reasoning. Fortunately, the reasoner helps in building these definitions by performing subsumption and consistency checks. The lowest level, primitive class is Building. It holds all the asserted buildings from the topographic dataset. The next higher-level classes in our abstraction hierarchy are those describing the types of dwelling. Their definitions are based on what we know intuitively (see chapter 3) and what can be found in any dictionary. For example, a semi-detached house is defined in the Oxford English Dictionary as "designating either of a pair of houses joined together and forming a block by themselves". To distinguish the many different types of buildings, we must consider both the immediate neighbourhood (e.g. touch relations) and the size of the building as an indicator of its purpose. To model these criteria, we require both object and data type properties of OWL. The object property 'touches' is created and made symmetric. This means if individual a touches individual b then individual b also touches individual a. The data type property 'hasArea' links an individual to the data type float to express a numerical decimal value. The data type is set to functional because a building can only hold one size value. This way, we can express both the calculated area value (in square metres) of the individuals (see ABox assertions), and we can constrain the class definition. We can now build our class definitions as follows:

House ≡ Building
       and hasArea **some** float[<= 160]
       and hasArea **some** float[>= 35]

HouseExtension ≡ Building
          and touches **some** House
          and hasArea **some** float [<= 35]

OutBuilding ≡ Building
        and not (touches **some** House)
        and hasArea **some** float[<= 35]

DetachedHouse ≡ House

      and not (touches **some** House)

SemiDetachedHouse ≡ House

      and not EndTerracedHouse

      and touches **max** 1 House

TerracedHouse ≡ EndTerracedHouse **or** MidTerracedHouse

MidTerracedHouse ≡ House

      and touches **min** 2 House

EndTerracedHouse ≡ House

      and not MidTerracedHouse

      and touches **some** MidTerracedHouse

The logical connectors (and, or, not), quantifiers (some, only) and cardinal restrictions (max, min, exactly, value) are shown in bold. Negation is used to make classes disjoint. This is to say that for example an individual of SemiDetachedHouse cannot be an individual of EndTerracedHouse. In the definition of TerracedHouse, it is important to differentiate between houses in the middle of the row and those at the end, because at the end of a row, the house only touches one other house. This definition is similar to SemiDetchedHouse. However, by having the class MidTerracedHouse, we can say that the EndTerracedHouse touches some MidTerracedHouse. Important is that the classification has to be performed sequentially because some definitions are built upon other concepts' individuals that need to be already classified and asserted. For example, to classify EndTerracedHouse we need to know which individuals are part of the class MidTerracedHouse. In addition, step by step reasoning reduces computational complexity.

The next higher-level aggregates define urban blocks and districts. Humans typically identify blocks because of their similarities and defined boundaries by roads. The goal is to classify urban blocks according to the type of housing they contain, and then to aggregate them into districts of the same kind of blocks that are directly connected.

Again, the neighbourhood of blocks plays a role in the recognition of their high-level meaning, i.e., districts. Therefore, to define the next set of classes, we introduce the object properties 'connectedTo' (symmetric) and 'contains', and the functional data type properties 'hasPercentageSemis', 'hasPercentageDetached' and 'hasPercentageTerraces'. With the -percentage- properties, it is possible to account for the degree of vagueness in the number of types of housing that are contained in one block. Hence, we can differentiate and define the following:

BlockTerrachedHouses ≡ hasPercentageTerraces **some** float[>= 70]

BlockSemiDetachedHouses ≡ hasPercentageSemis **some** float[>= 70]

BlockDetachedHouses ≡ hasPercentageDetached **some** float[>= 70]

BlockMixedHouses ≡ (contains **some** DetachedHouse
                  and contains **some** SemiDetachedHouse)
            or (contains **some** DetachedHouse
                  and contains **some** TerracedHouse)
            or (contains **some** SemiDetachedHouse
                  and contains **some** TerracedHouse)

DistrictTerracedHouses ≡ BlockTerracedHouses
                and connectedTo **some** BlockTerracedHouses

DistrictSemiDetachedHouses ≡ BlockSemiDetachedHouses
                  and connectedTo **some** BlockSemiDetachedHouses

DistrictDetachedHouses ≡ BlockDetachedHouses
                and connectedTo **some** BlockDetachedHouses

DistrictMixedHouses ≡ BlockMixedHouses
              and connectedTo **some** BlockMixedHouses

The final class, the one we wish to make explicit within the topographic dataset, is residential area. The definition of residential area is now straightforward because the

residential area covers the defined districts above. This can be expressed with a simple covering axiom such that:

ResidentialArea ≡ DistrictMixedHouses **or** DistrictDetachedHouses **or** DistrictSemiDetachedHouses **or** DistrictTerracedHouses

All these definitions are applied in section 7.2 for the concept-based instance retrieval, and are explained in more detail where necessary.

The ABox

The ABox is the extensional component of a DL system and represents the actual database or information store in terms of so-called assertions. The ABox is extracted from the data in the database and contains closed ground formulas, also called facts (Esposito *et al.*, 2007). Because the terminological knowledge is defined at an abstract logical level, data features stored in the spatial database must also be asserted in symbolic form to enable reasoning over them. The necessary knowledge is computed from the concrete geometry of the map to represent as many spatial aspects as possible. Due to the symbolic form of the DL, we can only represent qualitative spatial relationships. However, we can compute these spatial relations (based on the so-called region connection calculus (RCC)) from the geometry of the map, and represent these by means of RCC role assertions such as 'touches' and 'contains' in the ABox. This leads to a bottom-up computation of a potentially very large number of pair wise spatial relations, from which only a small number may play a part in the high-level interpretation (e.g. Neumann and Möller, 2008). Nevertheless, by integrating quantitative computations into the high-level concepts, a more efficient and transparent solution may be achieved. Selected spatial attributes such as area and length can also be represented using the concrete domain by means of data property assertions, e.g. hasArea = 12.34. This knowledge is necessary to recognise an individual in the ABox as an instance of one of the higher-level classes through size constraints and RCC appropriate role assertions. Furthermore, when querying classes that contain or imply a universal role or number restriction, we can answer queries completely only if we turn on closed domain reasoning mode. This means we have to close the ABox assertions with respect to the RCC role assertions. The presence of individuals in a knowledge base makes reasoning more complex from a computational viewpoint, as we will find

out in section 7.2 (Baader *et al.*, 2003). Nevertheless, given an ABox with concrete views as individuals, the DL system can generate an interpretation including all additional individuals that are required to satisfy the conceptual framework.

Linking the database and the knowledge base

Accessing external data sources that are independent from the ontology such as relational databases is problematic. Whereas databases are natural candidates for the management of the data layer, ontologies are the best candidates for realising the conceptual layer (Calvanese *et al.*, 2006a). The ontology is a virtual representation of a universe of discourse (i.e., domain), and the instances of concepts and roles in the ontology are simply an abstract representation of some real data stored in existing data sources. Unfortunately, most work on description logics do not deal with the problem of how to store ABox assertions or how to acquire these assertions from existing data sources. Therefore, establishing sound mechanisms for linking existing data to the instances of concepts and the roles in the ontology is of special importance wherever the use of ontologies is advocated. The mapping between relational data and ontologies is an important research topic, for example explored by Calvanese *et al.* (2006a and 2007b), Poggi *et al.*, (2008), and Dolbear and Goodwin (2007). To integrate external data sources with an ontology, we have to deal with the so-called impedance mismatch problem. This problem arises from the difference between the basic elements managed by the sources, namely data, and the elements managed by the ontology, namely abstract objects (Poggi *et al.*, 2008).

There are different ways to overcome this problem. For example, Protégé plug-ins such as the spreadsheet importer (Kola and Rector, 2007) and DataMaster (Nyulas *et al.*, 2007) allow you to import data from relational databases into ontologies. With DataMaster you can connect directly to a ODBC data source and import the data as classes and instances into the ontology. This creates a mapping between the database schema structure and ontology concepts. Table contents are imported as instances of the created table name class. These kinds of plug-ins represent an important part of the semantic data-access layer, which annotates and integrates disparate data sources into a semantically uniform data stream (Nyulas *et al.*, 2007). However, a major persisting limitation is the derivation of ontologies with flat structures that simply mirror the schema of the source databases (Cerbah, 2008). For the purpose of this thesis, I use a

custom approach where the database information is translated first into RDF, which then easily implements as OWL syntax in Protégé. Figure 42 illustrates the overall process.

```
┌─────────────────────┐
│    KB (Protégé)     │
└─────────────────────┘
          ▲
┌─────────────────────┐
│   Add RDF syntax    │
│    to OWL file      │
└─────────────────────┘
          ▲
┌─────────────────────┐
│   Python Script     │
└─────────────────────┘
          ▲
┌─────────────────────┐
│   Export CSV file   │
└─────────────────────┘
          ▲
┌─────────────────────┐
│  SQL queries / data │
│     preparation     │
└─────────────────────┘
          ▲
┌─────────────────────┐
│    DB (Oracle)      │
└─────────────────────┘
```
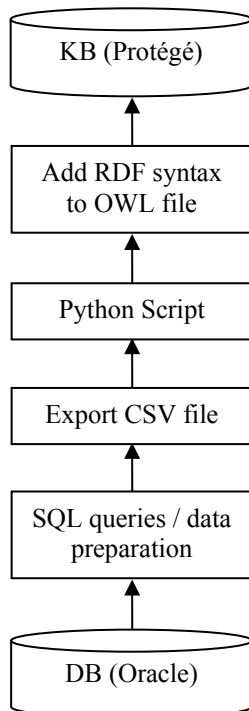
Figure 42 Translating database instances into OWL individuals

Initially, we start out with the information explicitly stored in the database. OS MasterMap, for example, is delivered in GML (geographic mark-up language) format. The data was converted and partitioned into Oracle tables according to the OS data model using Snowflake's GoLoader software. We then have to extract a suitable sample area (see chapter 4 for the overall knowledge discovery process). For the purpose of this proof of concept, I extracted three small samples from the Glasgow area using the SDO_WITHIN_DISTANCE operator set to a 1000 metres radius. The SQL syntax for the necessary operations is given in appendix D. From the reduced datasets, I then extracted only buildings to reduce the number of individuals. This will help to simplify the reasoning inside the DL system. Using the table consisting of only buildings, we can then calculate the touch relations between all buildings using the SDO_TOUCH operator. By joining the original BUILDINGS table with the derived TOUCH table, we create one table with all the required information for export into a comma separated file (CSV). The file contains the information from the following relevant columns: TOID, OSMM DESCRIPTIVE GROUP, CALCULATEDAREAVALUE, NUMBER_OF_BUILDINGS (that the object touches), and TOID_BUILDING2,

TOID_BUILDING3, etc. (TOIDs of the buildings that the first TOID touches). The unique topographic identifier (TOID) will serve also as an identifier for the asserted individuals in the ABox, which is important to be able to integrate the inferred results back into the database.

For deriving the type of urban blocks and districts, we do the same by extracting a text file of the urban block partitions, the blocks that touch one another, the individual buildings each block contains, and the percentage of the type of housing they contain (based on the prior inference of type of housing in Protégé). These operations were carried out in Radius Clarity 2.6 from 1Spatial. The algorithm for partitioning the vector dataset has been developed in house by the Generalisation Team at Ordnance Survey Research and was kindly provided for this work.

With a python script we can then translate the database information contained in the CSV files into RDF. The python script contains RDF syntax for describing an individual of the ABox. By creating an example individual within Protégé in the way we want all individuals to be asserted, we generate the necessary OWL code that the RDF syntax in the Python script must reflect. Instead of the details of that one individual, we assign which row of the CSV file will provide the information to be populated into the syntax (see appendix D for the code). The script then imports and runs through the individual CSV files, and populates the RDF syntax with the information stored in each row of that file. The output is the required syntax of all individuals for the OWL file. After copying and pasting the generated syntax into the OWL file, we can load the OWL file into Protégé, which then contains all the asserted data individuals. Figure 43 shows the loaded assertions for the building individuals.
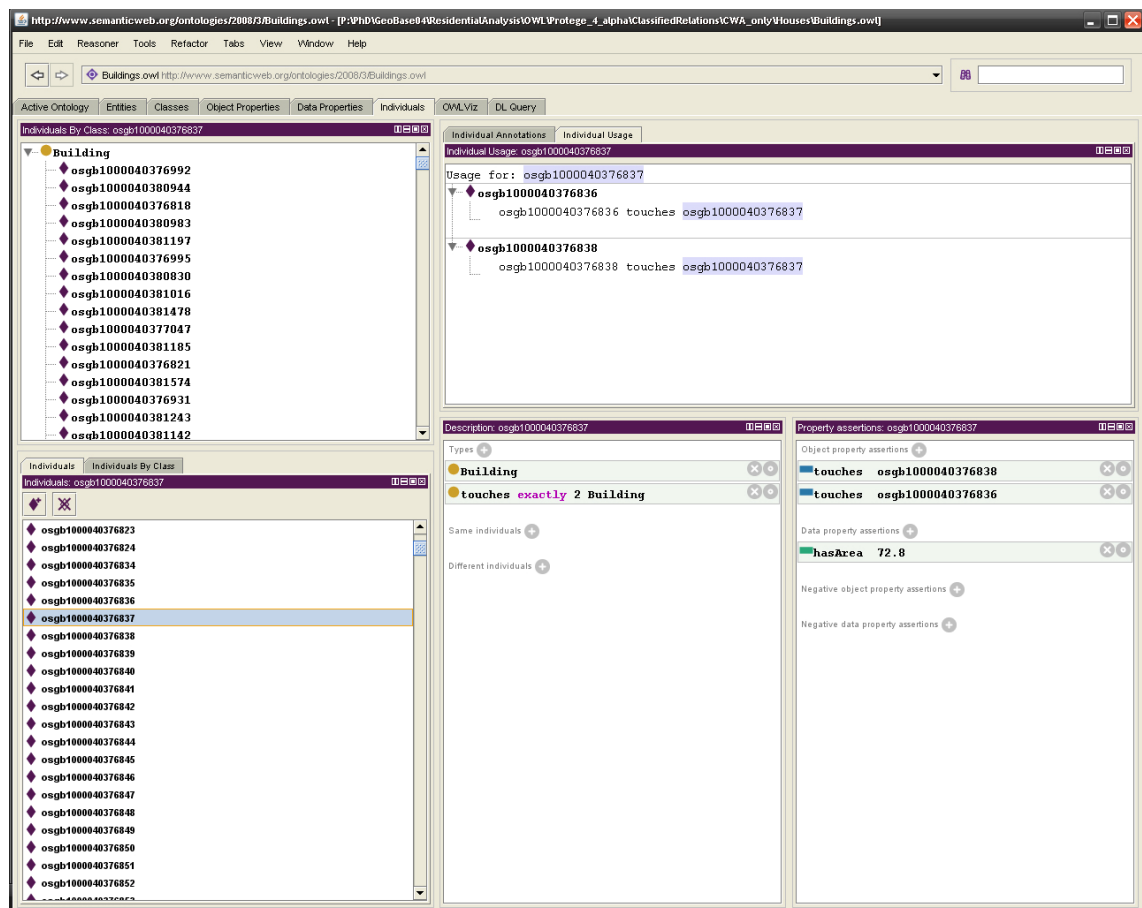
Figure 43 Asserted building individuals in Protégé

The knowledge that is asserted is the minimum required for making the inferences within Protégé. Relational databases store only values, therefore objects that are instances of the concepts in the knowledge base need to be constructed from such values. For the building individuals, we need to know how many other buildings a building individual touches, which ones it exactly touches, and its size in terms of the calculated area value (in square metres). Say we have an individual defined as follows:

- OSGB1000040381257 is a type of Building
- touches exactly 1 Building
- touches OSGB1000040381258
- hasArea 73.4

The data property 'hasArea' relates the individual to the value of $73.4m^2$. Values are external mathematical abstractions. Logic provides no function for calculating values; it can only conclude new statements from existing ones. That is why a mapping to databases is so important. Calculating the touch relation between buildings is also necessary. It is not enough to just say how many buildings a building touches, such as a

detached house touches no other building, a terraced house touches at least two buildings, a semi-detached touches exactly one other building. This would lead to misclassifications as not all buildings are houses. For example, a house may touch an extension (e.g. a conservatory) but no other house, meaning it is a detached house. However, based on only the knowledge that it touches another building, it will be classified as semi-detached house. That is why we need to set criteria for what is a house, and we need to know whether the building that the building is touching is a house or a house extension. Consequently, it is essential to model relations between individuals, such as knowing the other building's TOID.

For the urban blocks, we need to assert for example:

- UB63991 is an individual
- contains only {OSGB1000040377135, OSGB1000040377166, …}
- contains OSGB1000040377135, contains OSGB1000040377166, contains …
- connectedTo UB66244, connectedTo …, …
- hasPercentageDetached 0
- hasPercentageSemis 80
- hasPercentageTerraces 20

Each individual has a unique name that allows us to link it back to the database, in this case an identifier for the partitioned blocks. The role 'contains' is a simple object property that relates block individuals to building individuals. Therefore, we can say for each urban block which building TOIDs it contains. By linking these two types of individuals, we can then later classify urban blocks according to the type of housing they contain. It is important to close our individuals with the universal value restriction $\forall R.C$ (e.g., contains only). This means no other values exist except for those entailed by the axiom. Although we state which building individuals are contained in an urban block, the system does not know if there are any not explicitly stated building individuals in the open world. This is because OWL automatically assumes the open-world condition where any model means success without consideration of missing evidence (Neumann and Weiss, 2003). Instead, we need to apply the closure axiom to facilitate a finite model based on the closed world assumption. Therefore, we need to explicitly state both that an urban block contains a building TOID and that it only contains that building TOID.

In addition, we need to assert the percentage of type of housing an individual block contains. These values are computed from the previously inferred types of housing in Protégé. Unfortunately, OWL does not incorporate fuzzy/probability logic, which would allow us to express imprecise or vague knowledge. Equally, we cannot count individuals inside a knowledge base to derive percentage values. Consequently, this knowledge needs to be calculated outside the DL system and then asserted explicitly to enable reasoning over percentages. For this, we specify three properties that carry the different house type percentages respectively.

Lastly, we assert which block individuals touch one another. Similar to describing which buildings touch, the knowledge base needs to know which blocks are connected to be able to infer individuals of the district classes. From this, we can establish which blocks of the same kind are connected to one another. The described assertions for urban block individuals are shown in figure 44. The OWL code of an asserted urban block individual is given in appendix D. Next, we use both the defined classes and asserted individuals to infer and classify the instances of the higher-level concepts.
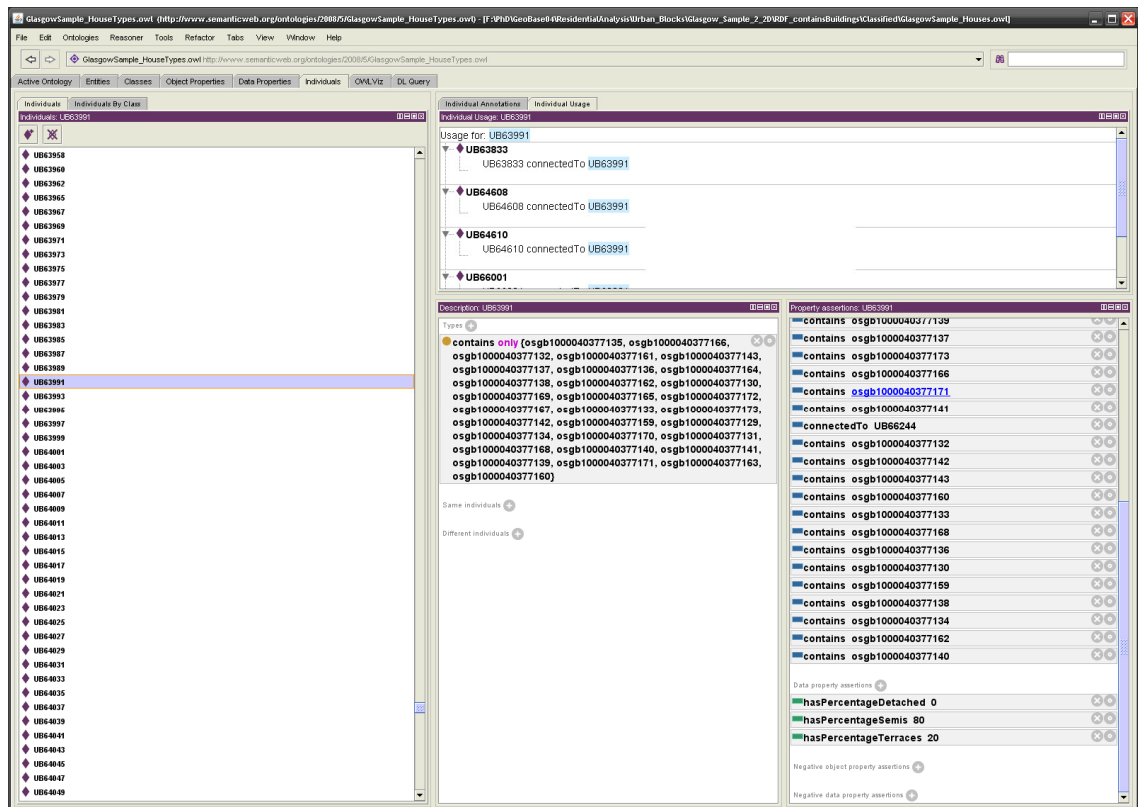
Figure 44 Asserted urban block individuals in Protégé

7.2 Concept-based instance retrieval and classification

Query answering with respect to an ontology is in general a deductive process of finding domain objects that satisfy the query in all possible worlds constrained by the ontology (Calvanese *et al.*, 2007c). For instance, by storing a basic set of relationships from the domain, a logic-based system deduces others from the basic ones if it needs them in answering a query (Rips, 1994). ABox query answering in particular is used to implement retrieval systems based on high-level interpretations of data objects. The user poses a query that describes the information he or she wants to retrieve in terms of the underlying terminology of the ontology. Since the ontology is conceptually close to the high-level vocabulary of the user, queries appear intuitively more natural.

The success indicators of query answering are CPU time for answering the query and the amount of memory used. This relates directly to the problem of DL expressivity and deductive complexity. To rate the efficiency with respect to scalability, you have to take into account the size and complexity of the source ABox. Experience with ontologies derived from database content has shown that it is often necessary to effectively solve instance retrieval problems with respect to huge amounts of data descriptions that make up major parts of ontologies (Haarslev and Möller, 2008 and 2001). Although reasoners such as FACT++ and PELLET are based on a tableau reasoning algorithm and integrate various optimisation techniques to provide for a fast and efficient practical implementation (Esposito *et al.*, 2007), the size of the ABox and the complexity of the assertions ultimately determines the speed of query answering.

The following sections outline the high-level inferences of our asserted knowledge base. Each section respectively describes the inference according to every stage of the abstraction hierarchy of our conceptual model (chapter 5). Unfortunately, current DL reasoning systems do not yet provide the services that would optimally support high-level inferences since concrete views do not provide logically sufficient conditions for higher-level classification. In other words, logic does not provide the necessary means to autonomously derive all the information required for the inference. The problem of missing information as well as the complexities of reasoning are addressed for each classification level. The results are visualised by exporting the inferred knowledge back into the database and creating a thematic map of the classified instances.

Inference of type of dwelling

Housing type is a key variable in defining urban structures. Houses are typically classified into five types of purpose-built flats, converted flats, terraced, detached/semi-detached houses and miscellaneous buildings (Batty and Longley, 1994). The objective here is the detection and assignment of the dwelling types semi-detached, detached and terraced houses from building data. Figure 45 shows the ontology for classifying buildings. The ontology comprises primitive concepts (yellow) and defined concepts (orange), as discussed in chapter six. Whereas a primitive concept has no definition, a defined class means that its concept name is equivalent to its given definition.
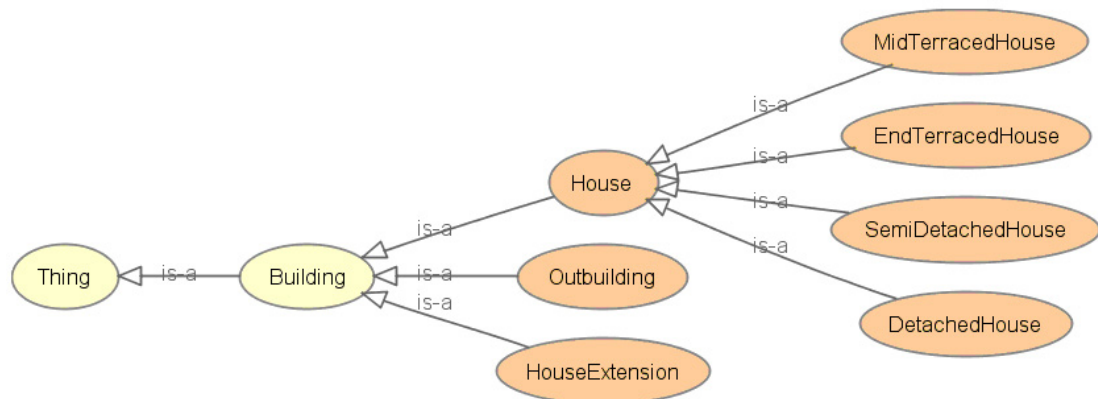
Figure 45 Dwelling taxonomy

Based on class definitions in the TBox and the asserted knowledge in the ABox, a DL reasoner infers which individuals are members of the respective classes. Because logic-based ontologies function on the notion of set theory, a concept is treated as a set of a well-defined collection of instances. In other words, the concept is defined in such a way that a DL reasoner can determine whether any given individual belongs to that set. If a class subsumes another class, then the individuals from the one class form a subset of the set of individuals from the other class. Therefore, it is important that classes are not made disjoint along the subsumption hierarchy because two disjoint classes cannot share the same set of individuals. For example, the class House cannot be disjoint from the class DetachedHouse because the subsumer shares a subset of its set of individuals. Otherwise, this would lead to an inconsistent knowledge base. However, classes on the same level of the hierarchy must be disjoint to ensure that individuals cannot be instances of more than one class. For example, instances of SemiDetachedHouse cannot be instances of DetachedHouse or EndTerracedHouse at the same time.

To get the correct definition of the classes, a very small dataset is used initially to tweak the defined classes. Figure 46 shows the different classified building types according to the ontology in figure 45. There is enough expressivity to discern between semi-detached, detached, mid-terraced and end-terraced houses, outbuildings (e.g. garage) and house extensions (such as conservatories). The OWL code for this small ontology and its asserted individuals can be found in appendix D.

Classified Building Types

🟥	DetachedHouse	(2)
🟩	EndTerracedHouse	(2)
🟨	HouseExtension	(11)
🟩	MidTerracedHouse	(12)
🟪	Outbuilding	(2)
🟦	SemiDetachedHouse	(2)
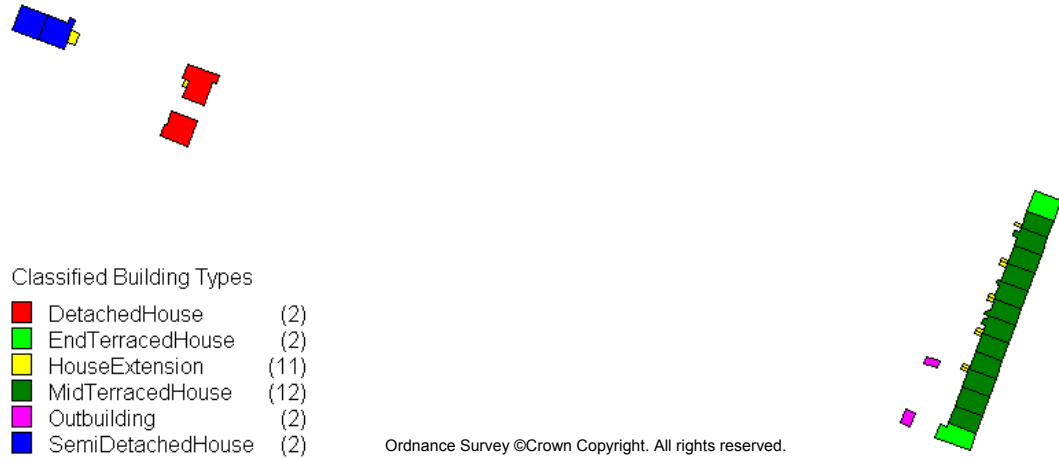
Figure 46 Illustrating expressivity of OWL to infer types of buildings

The main drawback of this fine-grained ontology is its computational complexity. The expressivity of the above definitions is $\mathcal{ALCIQ(D)}$, which includes concept intersection, universal and existential qualifiers, complex concept negation, cardinality restrictions, and symmetric and data type properties (see appendix C). The main issue is the size of the ABox and the number of complex relations asserted in the ABox. Large numbers of individuals that link to one another through the 'touches' property increase the number of relations that the reasoner must consider. This may work fine for a small sample as above, but when definitions are scaled up to a larger sample area, we run into computational problems. Since the inference is carried out in memory, the available memory runs quickly out when reasoning over large ABoxes.

Domain modelling and the inference therefore require complexity reduction. To improve computational efficiency for a large ABox, we need to simplify the ABox. Instead of importing all buildings of all sizes and letting the reasoner identify outbuildings and house extensions, we exclude small buildings (of size less than 35m^2)

prior to the classification. This step is done within the database using a SQL query to import only the buildings required for the reasoning inside the knowledge base. This reduces the number of relations that the reasoner must handle. For example, a terraced house now only touches other houses and not also small buildings such as a house extension. Hence, the number of relations is reduced significantly from four or five to a maximum of just two related building individuals. In addition, this allows us to simplify the ontology by dropping the classes Outbuilding and HouseExtension. Now we can define the dwelling terminology as follows:

DetachedHouse ≡ House
                        and not (touches **some** Building)

SemiDetachedHouse ≡ House
                        and not EndTerracedHouse
                        and touches **max** 1 Building

MidTerracedHouse ≡ House
                        and touches **min** 2 Building

EndTerracedHouse ≡ House
                        and not MidTerracedHouse
                        and touches **some** MidTerracedHouse

The definitions now assert the touch relation between individuals of the primitive class Building instead of the defined class House. This reduces the computational effort required by the reasoner. However, we still need to know which buildings some individuals touch, because EndTerracedHouse remains defined as a House that touches some MidTerracedHouse. Thus, for an individual to become a member of EndTerracedHouse, it needs to know that the building it touches has been classified as a MidTerracedHouse. Figure 47 shows the classification results based on the above definitions for a slightly larger sample area (individual count 503). The buildings that were excluded from the classification are also shown. Large buildings are not classified because they do not satisfy the size constraint of the class House, which the defined dwelling concepts subsume. We can see that the reasoner correctly inferred all instances of the defined classes.

Buildings classified by House Type
■ DetachedHouse (72)
■ EndTerracedHouse (64)
■ MidTerracedHouse (313)
■ SemiDetachedHouse (38)

Figure 47 Thematic map of inferred types of dwelling

Lastly, the ontology is applied to the three study areas. Giffnock consists of 4288 individuals of the type Building. Drumchapel has 4,314 buildings and Pollokshields has 2,941 buildings. These numbers exclude buildings of size less than $35m^2$. Figure 48 shows the Protégé interface after the reasoner classified all the instances in the ontology. We can see the defined class EndTerracedHouse and its inferred instances with a yellow backdrop. Figures 49 to 51 present the results of the classification in a thematic map of the three study areas, respectively. For all three study areas the same ontology rules were applied. However, in the case of Pollokshields, which consists largely of large Victorian villas, the size threshold of 160 m^2 was not very successful. Many of the detached dwelling were omitted by the reasoner because they are larger than $160m^2$. As the reconstruction cost table in table 8 has shown, different types of houses have different sizes depending also on their age. Villas are not listed in this table, however, they are much larger than ordinary residential detached houses. To accommodate for this effect, another classification was carried out with the size threshold of the class

House increased to 280m^2 (figure 52). Now the reasoner picks up all the Victorian villas. However, as the validation next will show, the increased threshold has other implications, such as misclassifying other non-residential, large buildings. This promotes the need for adapting the ontology rules to different types of houses instead of having one class defining house which subsumes the more specialised housing types.
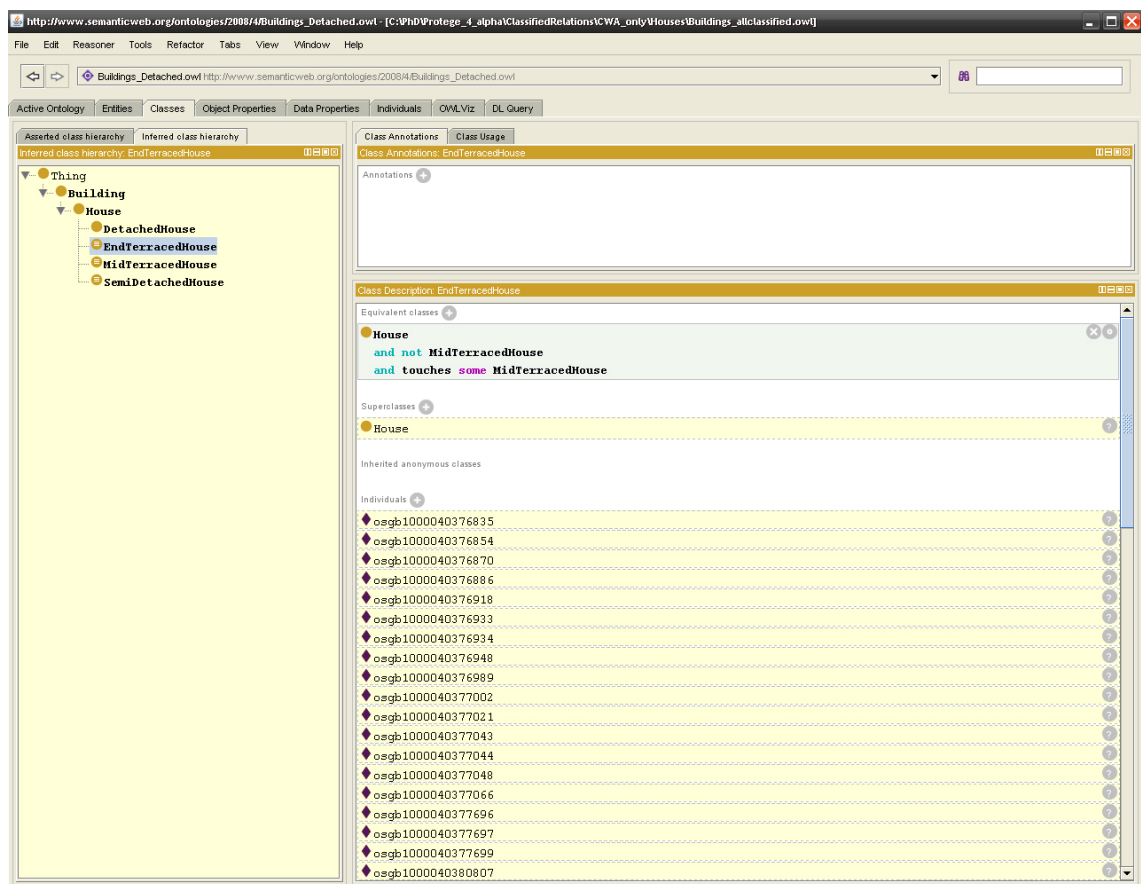
Figure 48 Inferred individuals in Protégé

Giffnock buildings classified by type of dwelling

■	DetachedHouse	(940)
■	EndTerracedHouse	(333)
■	MidTerracedHouse	(944)
■	SemiDetachedHouse	(1769)
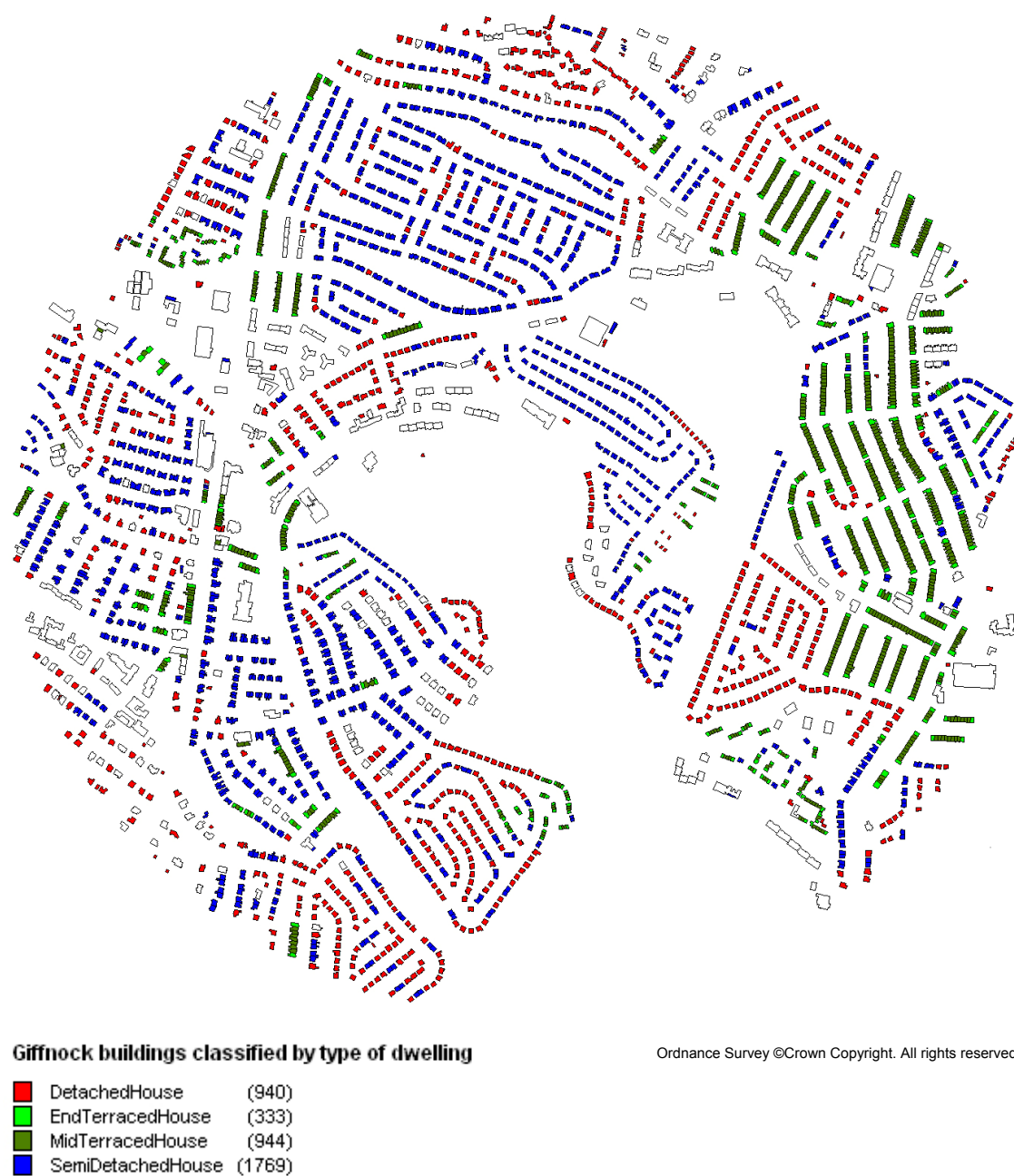
Figure 49 Giffnock classified by type of housing

Drumchapel buildings classified by type of dwelling

	DetachedHouse	(400)
	EndTerracedHouse	(383)
	MidTerracedHouse	(692)
	SemiDetachedHouse	(2615)

Figure 50 Drumchapel classified by type of housing

Pollokshields buildings classified by type of dwelling

🟥	DetachedHouse	(593)
🟩	EndTerracedHouse	(188)
🟩	MidTerracedHouse	(620)
🟦	SemiDetachedHouse	(544)

Figure 51 Pollokshields classified by type of housing

Pollokshields buildings classified by type of dwelling
Size threshold set to 280 square metres

■ DetachedHouse   (951)
■ EndTerracedHouse   (252)
■ MidTerracedHouse   (784)
■ SemiDetachedHouse (664)

Figure 52 Pollokshields classified by type of housing with increased threshold

To get a better understanding of the success of the classification, we can compare the results with Google Earth imagery and validate the number of residential buildings by using OS MasterMap Address Layer 2. OS MasterMap Address Layer is the most complete, comprehensive, national spatial address dataset for the whole of Great Britain. The information has been assembled from data collected by Ordnance Survey and from key organisations involved with the creation of addresses, notably Royal Mail and Valuation Office Agency (Ordnance Survey, 2006). Each address is classified as

either residential or commercial. Commercial address are further broken down where the trading or brand name provides clear details of their function, for example, B&Q equals retail. The dataset also includes buildings that may be known by a name as well as a house number. This includes sub-building names such Flat 1. Furthermore, a non-postal theme contains miscellaneous premises like churches, halls, car parks, and public conveniences. We can use this information for identifying residential buildings, in particular which ones are multi-occupancy, i.e., flats, and commercial buildings. This reference dataset will allow us to establish how successful the classification picked up residential houses and which buildings were misclassified.

In addition, we can do a visual inspection of the classified types of dwellings by transforming the shape file with the classified polygons into a KML file so that it can overlaid on top of Google Earth. For this purpose, I used the freely available tool Shp2kml[3]. Shp2kml is a stand alone tool that transforms GIS layers to Google Earth. Figure 53 shows all three classifications in Google Earth, giving an overview of the location of the study areas around Glasgow. Figures 54 to 56 show enlarged examples within the three study area, respectively. Although the projection is slightly off, we can still see that the classified polygons (blue outline for semi-detached houses, red for detached houses, light green for end-terraced houses and dark green outline for mid-terraced houses) matches the underlying aerial imagery of the real world buildings. In figure 54, we can see successfully classified semi-detached, terraced and detached houses in Giffnock. Figure 55 shows amongst others some misclassified semi-detached houses in Drumchapel. Lastly, in figure 56, we can see that Pollokshields' tenements were misclassified as terraced houses due to the increased size threshold. In reality, these rows of houses are multi-family dwellings as can be seen from OS MasterMap Address Layer in figure 59. Flats and non-residential addresses were added to the classified datasets of all three case studies (figures 57-59), thus highlighting which buildings were misclassified. With the help of these two methods, we can now look at some classification statistics and identify common errors in the classifications. In particular, we can create prediction success tables of all three case studies. Such confusion matrices give the prediction success as well as the proportions of objects that were actually classified.

---

[3] Zonums Solutions, available from URL: http://www.zonums.com/shp2kml.html

Figure 53 Aerial view of all three classified study areas

Figure 54 Giffnock – validation of classification in Google Earth

Figure 55 Drumchapel – validation of classification in Google Earth

Figure 56 Pollokshields – validation of classification in Google Earth
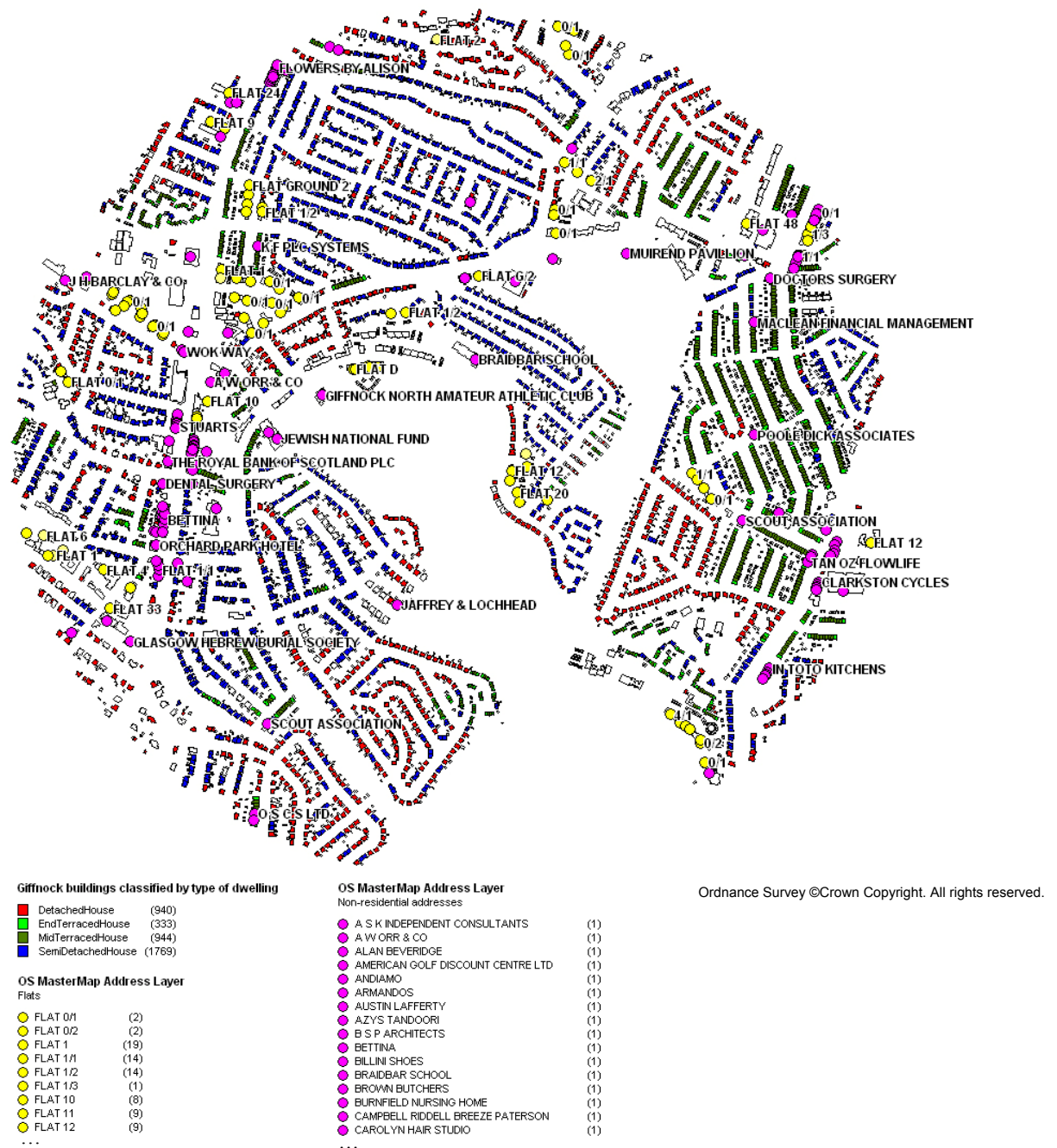
Figure 57 Giffnock OS MasterMap Address Layer with flats and non-residential addresses
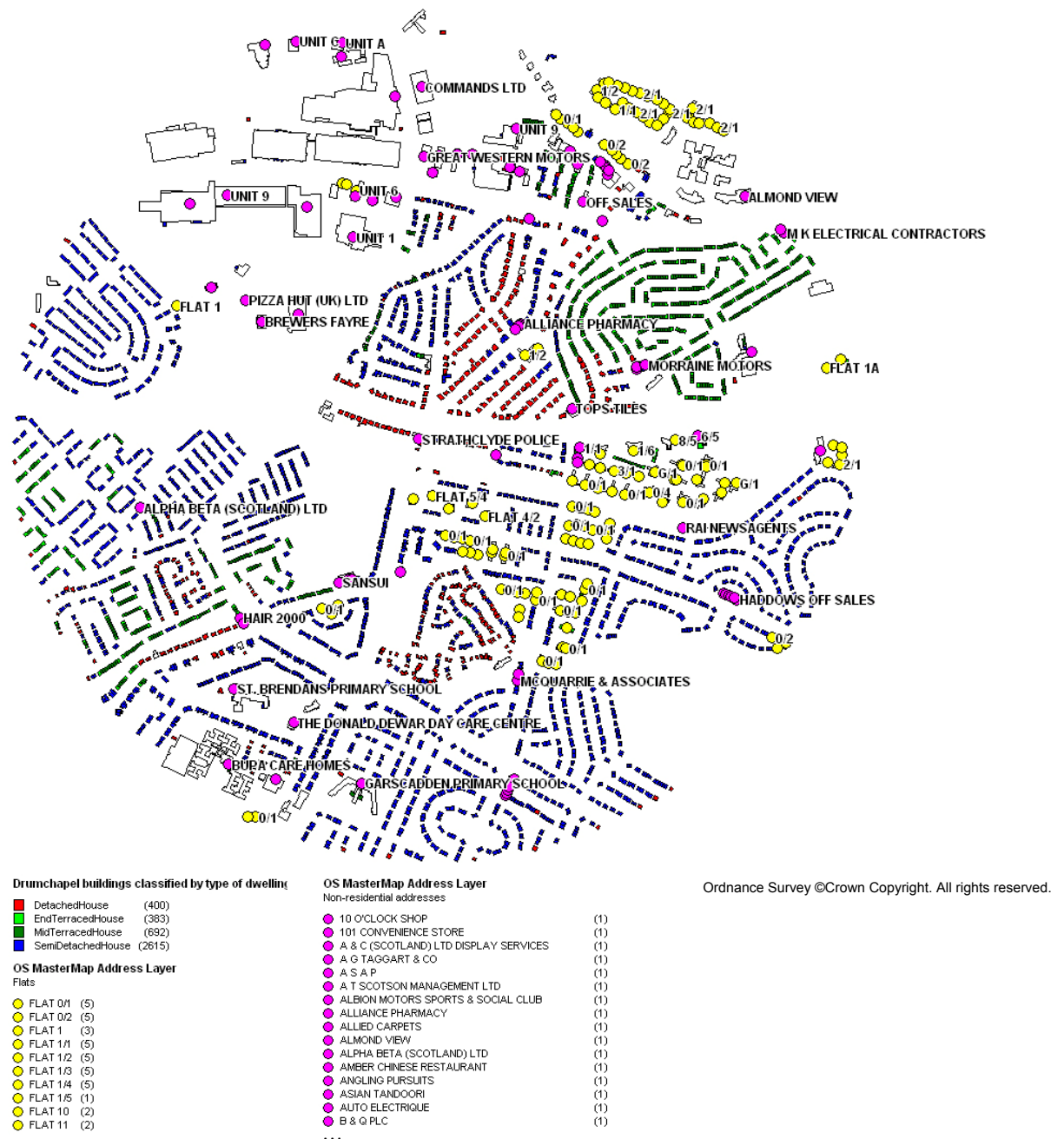
Figure 58 Drumchapel OS MasterMap Address Layer with flats and non-residential addresses

Figure 59 Pollokshields OS MasterMap Address Layer with flats and non-residential addresses

From OS MasterMap Address Layer 2 and the classified datasets, we can establish some building statistics about the three study areas. Table 9 summarises these by the total number of buildings, number of buildings greater than $35m^2$, number of buildings classified with the size threshold set to $160m^2$ and $280m^2$ (for Pollokshields), and the number of buildings with flats and non-residential buildings in each category for each study area. With OS MasterMap Address Layer overlaid on top of our datasets, we can query which buildings contain the address layer's point features where the sub-building attribute lists a flat and where the postal code organisation attribute lists a business or

other non-residential address, respectively. From this information we can establish the number of buildings in the original dataset as well as the classified dataset that are in fact multi-family dwellings or non-residential buildings. Overall, the areas are predominantly residential, however a small number of non-residential premises remain. For instance, looking at the classified datasets, there are 0.7% classified buildings with flats and 0.9% classified buildings which are non-residential in Drumchapel. In Giffnock, there are 0.3% classified buildings with flats and 2.1% classified buildings with non-residential addresses. Pollokshields has the largest number of flats with 4.8% classified buildings actually being multi-family dwellings and 1.5% classified buildings being non-residential. These percentages increase dramatically with the size threshold raised to 280m^2 in the Pollokshields' dataset. 13.7% of the classified buildings now include flats and 2.4% are non-residential.

Table 9 Building statistics from the classification and OSMM Address Layer 2

		Drumchapel	Giffnock	Pollokshields
Buildings	All buildings	6,004	8,454	5,852
	Buildings >35	4,314	4,288	2,941
	Classified <160	4,090	3,986	1,945
	Classified <280	-	-	2,651
Flats	All buildings	153	101	549
	Buildings >35	153	101	548
	Classified <160	29	15	94
	Classified <280	-	-	365
Non-residential	All buildings	83	127	114
	Buildings >35	82	124	112
	Classified <160	40	85	29
	Classified <280	-	-	63

On first impression, the reasoner fairly accurately inferred the different types of houses with the minimal knowledge that has been asserted in the ABox, such as an individual is a building, it has a certain size, and knowing how many and which buildings it touches. However, on closer inspection we will find common errors in all three datasets. For example, figure 60 illustrates typical misclassifications (left) and omissions (right) of detached dwellings. The result is shown in the context of the complete OS MasterMap dataset. Some outbuildings near terraced houses were misclassified as detached houses, and some detached houses were omitted because of the threshold value in the definition of the class House (less than 160m^2). Since detached houses are generally larger in size than semi-detached and terraced houses (see table 8), the class DetachedHouse suffered the most omissions. However, if you increase the threshold, as done in the case of Pollokshields, then the reasoner classifies buildings as house that are not actually

residential. Here it would be worth while to adapt the rules to account for different house sizes. In addition, the parameters could be improved by running some statistical analysis for finding the most appropriate threshold for a given area. For this purpose, pattern recognition and computer visions techniques could be used for describing shape, proximity and configuration statistics. Such methods allow us to determine the correlation between many potentially geometric and attribute factors and the required classification. Indeed, there will always be exceptions to the rule. This is generally an issue when working with parameters and defined threshold – whether in a knowledge base or in a programmed algorithm.
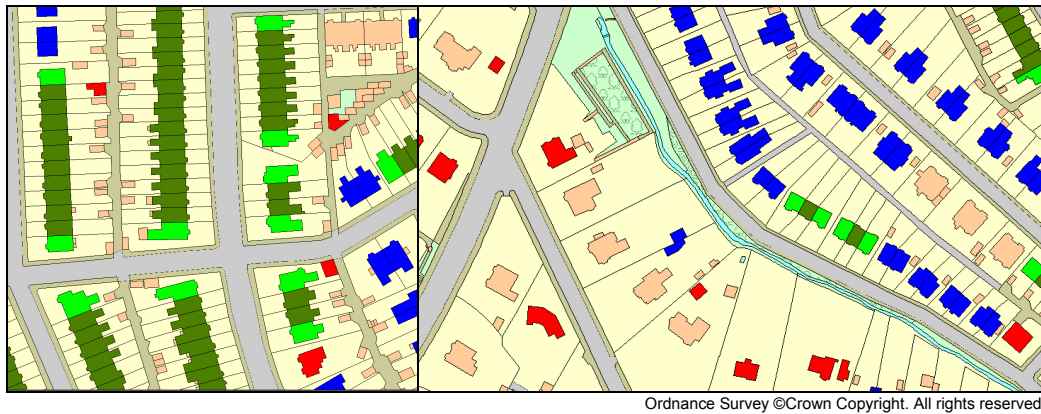
Figure 60 Typical misclassification and omission of detached dwellings

The threshold value also caused some omissions and misclassifications in the class SemiDetachedHouse (figure 61). Misclassifications include buildings where two buildings are adjacent to one another, but one is much larger than the threshold value. Alternatively, two buildings may be touching, but in the context of the scene, they are more likely to be a pair of garages or outbuildings. The definitions could be refined by stating that both houses in a pair must be of similar size and shape to be classified as SemiDetachedHouse. The questionnaire survey in chapter 3 showed the importance of additional factors such as shape, proximity and orientation (see figure 15). For example, we could compute the shape of a polygon area by the squared hull and derive absolute orientation by looking at the longest edge of a polygon (Steiniger and Weibel, 2005). Equally, we can include more context information. The survey also captured a lot of knowledge about the residential land use domain. As table 6 summarises from the questionnaire, a house is defined as a building next to a garden. OS MasterMap, for instance, describes residential gardens as 'multi-surface'. This additional knowledge

could be asserted in the knowledge base to improve the classification of all residential dwelling types.
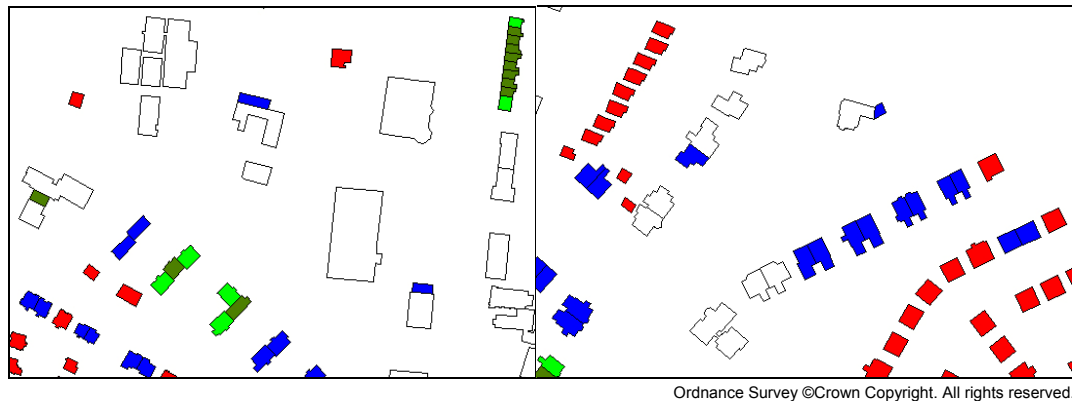
Figure 61 Typical misclassification and omission of semi-detached dwellings

In the case of terraced houses, misclassifications occurred in all three datasets because we simplified the concept definitions of our ontology. By reducing the definition of MidTerracedHouse to a House that touches exactly two Building (instead of two House), different building combinations were classified as terraced houses. Whereas one building satisfies the size criteria of house and touches exactly two buildings, these buildings may not satisfy the house criteria in that they are too large. This is then not a scenario of a row of residential houses, but a group of connected buildings that serve an entirely different function, as shown in figure 62. Although this can be rectified by altering the definition, the computational effort to go through all the relations between large numbers of individuals is potentially too large. Some omissions also become visible in the context of the complete dataset (figure 63). The exclusion of small buildings from the classified datasets meant that some small terraced houses were excluded from the classification.
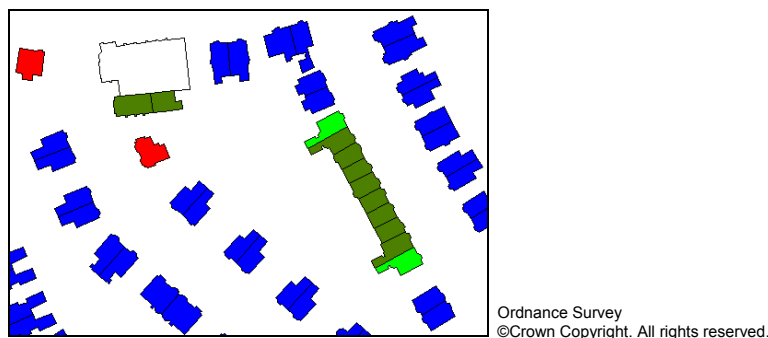
Figure 62 Misclassification of terraced dwellings
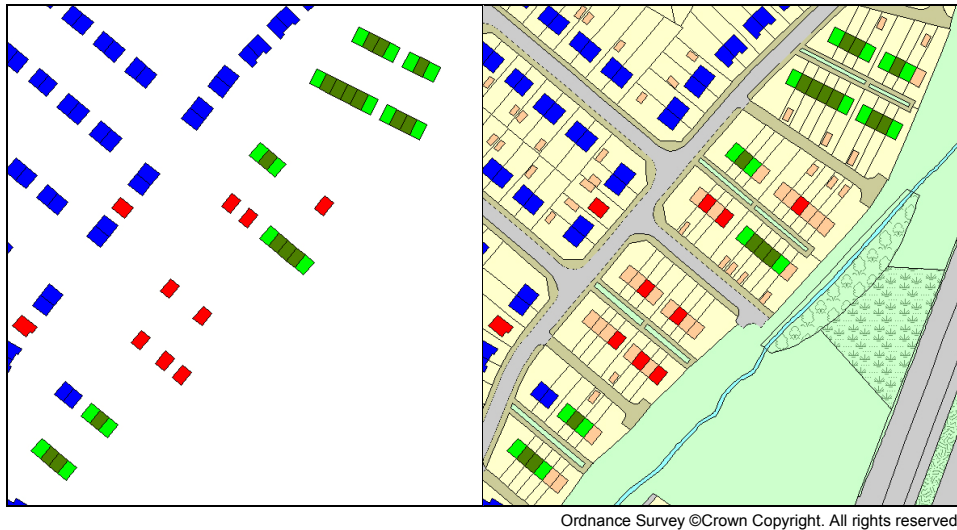
Figure 63 Omission of terraced dwellings

As we can see from the previous figures, common errors were caused by excluding small buildings from the classification. Other errors resulted from the lack of context information in the definitions. This however remains a predicament: Although results would be more accurate with more knowledge, the complexity would increase to such a degree that the reasoner potentially fails to compute a solution. Based on the applied rules in the ontology, tables 10 to 13 summarise the classification success of all three case studies including the one with the increased threshold. In all four classifications, the reasoner misclassified a lot of buildings which do not belong to any of the dwelling categories. These misclassification are grouped collectively under the category 'others'. This includes the misclassification of non-residential and multi-family buildings. In Giffnock and Drumchapel the reasoner only misclassified a small percentage of 'other' buildings, that is, 2.9% and 2.3% respectively. In Pollokshields on the other hand, the dataset contains large numbers of tenements, which resulted in 17.6% misclassification of 'other' buildings. With the size threshold raised to 280m^2, the misclassification increased to 23.4%. In terms of the classified categories detached, semi-detached, end-terraced and mid-terraced house, the reasoner did overall a good job. Especially in Pollokshields, there was not much confusion among the categories. In the Giffnock and Drumchapel datasets, the highest confusion is between detached and end-terraced houses. This happened because some of the terraced houses were omitted by having set the lower size limit to 35m^2. This meant that actual end-terraced houses became single houses. This was also the case for some mid-terraced and semi-detached houses. Equally, a few semi-detached houses were confused with mid- and end-terraced houses.

In the Pollokshields dataset, only the class detached house was confused with the class semi-detached house. This happened where detached houses touch a large garage or a house extension, which were also classified as house.

Table 10 Giffnock confusion matrix

		Actual					Total
		Detached	End-terraced	Mid-terraced	Semi-detached	Other	
Predicted	Detached	**860**	13	5	0	62	940
	Endterraced	0	**313**	9	1	10	333
	Midterraced	0	0	**932**	1	11	944
	Semi-detached	3	2	4	**1727**	33	1769
	Total	863	328	950	1729	116	**3,986**

Table 11 Drumchapel confusion matrix

		Actual					Total
		Detached	End-terraced	Mid-terraced	Semi-detached	Other	
Predicted	Detached	**305**	57	7	9	22	400
	Endterraced	0	**365**	2	2	14	383
	Midterraced	0	3	**656**	2	31	692
	Semi-detached	0	11	15	**2561**	28	2615
	Total	305	436	680	2574	95	**4,090**

Table 12 Pollokshields confusion matrix at same threshold as other study areas

		Actual					Total
		Detached	End-terraced	Mid-terraced	Semi-detached	Other	
Predicted	Detached	**436**	0	0	0	157	593
	Endterraced	0	**160**	0	0	28	188
	Midterraced	0	0	**538**	0	82	620
	Semi-detached	3	0	0	**466**	75	544
	Total	439	160	538	466	342	**1,945**

Table 13 Pollokshields confusion matrix at increased threshold (280m^2)

		Actual					Total
		Detached	End-terraced	Mid-terraced	Semi-detached	Other	
Predicted	Detached	772	0	0	0	179	951
	Endterraced	2	163	0	0	87	252
	Midterraced	0	0	542	0	242	784
	Semi-detached	14	0	0	537	113	664
	Total	788	163	542	537	621	**2,651**

Lastly, figures 64 to 66 illustrate the classifications in the context of OS MasterMap Topography Layer including all other non-building features. Despite some omissions, misclassifications and the general computational issues, the ontology effectively infers higher-level knowledge in terms of types of dwellings. Considering the level of fuzziness in higher levels of the model, the errors occurring at this level should be well within the fuzziness of the next higher-level class definitions. Therefore, the misclassifications and omissions should not pose any issues for a generalised view of residential area.

Giffnock buildings classified by type of dwelling

🟥	DetachedHouse	(940)
🟩	EndTerracedHouse	(333)
🟫	MidTerracedHouse	(944)
🟦	SemiDetachedHouse	(1769)

Figure 64 Giffnock classification visualised in OS MasterMap

Drumchapel buildings classified by type of dwelling

🟥	DetachedHouse	(400)
🟩	EndTerracedHouse	(383)
🟩	MidTerracedHouse	(692)
🟦	SemiDetachedHouse	(2615)
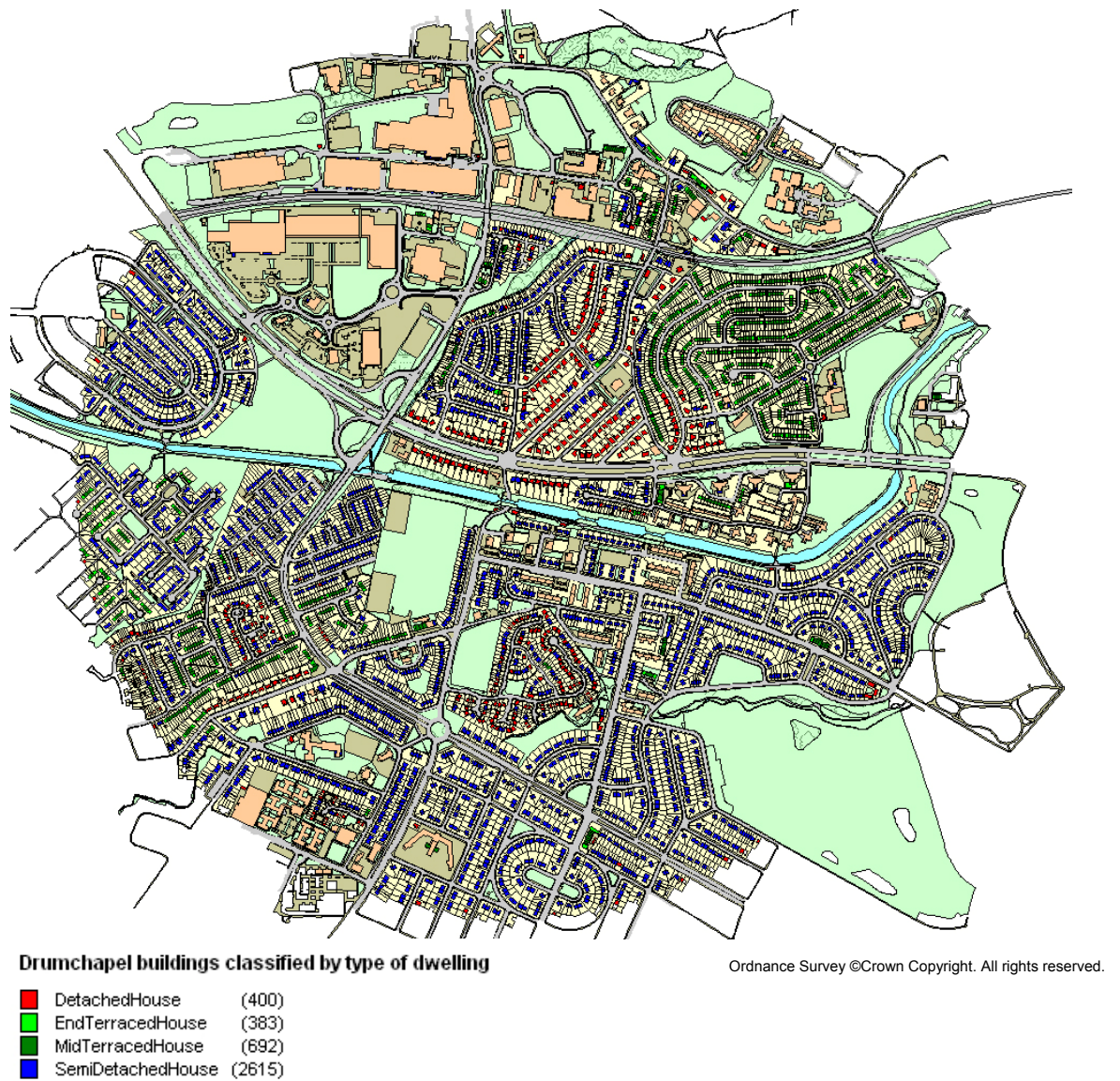
Figure 65 Drumchapel classification visualised in OS MasterMap

Pollokshields buildings classified by type of dwelling

- DetachedHouse (593)
- EndTerracedHouse (188)
- MidTerracedHouse (620)
- SemiDetachedHouse (544)

Figure 66 Pollokshields classification visualised in OS MasterMap

The beauty of this approach is that a DL system employs standard reasoning services that not only save on software development efforts but also provide explanations about their inferences. The definitions can be easily changed, or new concepts added. The whole process from modelling to inference is explicit. Subsequently, the inferred instances are asserted back into the knowledge base as individuals of their derived parent classes to assist in the classification of the next higher-level concepts. That way, the ABox part of the ontology is extended with new assertions describing individuals of types of dwellings, which are used for further data characterisation and classifying the next higher-level concepts of types of blocks and districts.

Inference of type of urban block

"The plots of private land surrounded by public streets are like an archipelago of islands set in a sea of public space."

–Steven Marshall (2005, p.13)

The urban block forms an integral part in neighbourhood models of urban morphology. It provides constraints for guiding the global partitioning of building sets on the whole map by means of roads and rivers (Li et al., 2004). These partitions form blocks, districts and neighbourhoods, which are defined through Gestalt principles such as similarity and proximity. As explained in chapter 5, the block forms an important aggregate structure for this conceptual model. However, this implementation requires auxiliary methods for calculating missing knowledge such as the urban blocks. The algorithm used for this purpose is based on the road network, partitioning islands between roads. The block itself is a connected surface delimited by looping streets that contain buildings (Larive et al., 2005; Gaffuri and Trévisan, 2004). The block is therefore a meso object that combines and generalises elementary features in relation to surrounding roads into groups (Ruas, 2000). It serves as a convenient unit for partitioning the space into more manageable chunks, mainly to reduce computational complexity for generalisation algorithms. Although deriving meso objects requires knowledge and human capacity of interpretation in interactive processes (Boffet, 2000), we have to adapt the divide and conquer strategy of a partitioning algorithm. Similar to the missing touch relations in the previous section, we have to compute the partitions of our sample datasets outside the DL system. However, with the expert knowledge held in the DL system, we can reason about the block partitions, classify them according to the type of housing they contain, and infer implicit knowledge.

The classification of blocks forms the second stage of this hierarchical interpretation. The reasoner classifies blocks according to their heterogeneity, that is, the type of housing each block contains. Based on the previously inferred types of dwelling, we can now classify individual blocks as members of the class BlockDetachedHouses, BlockSemiDetachedHouses, BlockTerracedHouses, or BlockMixedHouses. Figure 67 shows the ontology with the defined classes denoting the different types of blocks. From the dwelling taxonomy (figure 45) only the named classes for each type of dwelling remain to hold their asserted instances, respectively. The class TerracedHouse

remains defined as a covering axiom for the classes EndTerracedHouse and MidTerracedHouse. Because of the additional asserted knowledge, for example in the Giffnock ontology, we now have a class count of 16 and the number of individuals rises to 10,176. There are 8,404 class assertion axioms, 5,214 data property assertion axioms, and 17,229 object property assertion axioms. This illustrates how the amount of data we are dealing with increases the complexity of the DL system. The DL expressivity is $\mathcal{SOF(D)}$, denoting $\mathcal{ALC}$ as well as nominal and functional properties.
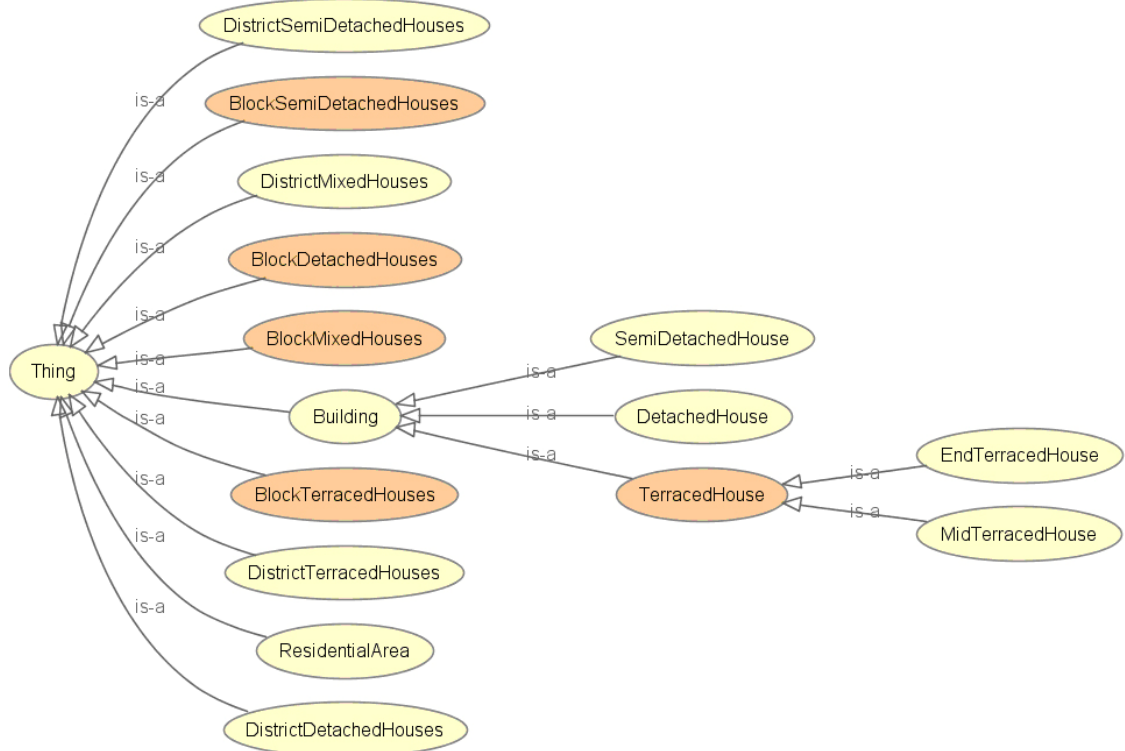
Figure 67 Taxonomy of defined urban block classes

We have knowledge about a given set of urban block individuals and the TOIDs they link to with the 'contains' property. We now want to infer which type of dwelling the blocks contain and classify them accordingly. We can do a 100% classification by defining that for example the class BlockDetachedHouses contains only DetachedHouse. However, reality is much more uncertain and vague. If the class BlockDetachedHouses contained only one single terraced house, it would be classified as BlockMixedHouses. Most blocks contain a mixture of types of housing, usually with a clear majority. Consequently, we need to classify blocks in relation to this majority to avoid having most of the blocks classified as being of type mixed houses. Ideally, we would want Protégé to tell us that instances of BlockDetachedHouses contain a majority

of detached houses (perhaps greater than 70%). Unfortunately, logic only allows one to make inferences about statements. As a result, we need to go outside the DL system again to calculate the missing knowledge.

These limitations can be potentially overcome in the future as logic is evolving to permit reasoning over probabilistic statements (Klinov, 2008). The probabilistic reasoner Pronto is an extension to Pellet and provides probabilistic reasoning over OWL-DL ontologies. Although Pronto is not yet implemented in Protégé, it processes an existing OWL ontology by adding to and inferring new probabilistic statements from a probabilistic ontology. Probabilistic reasoning refers to the probability of a class being a sub-class of another class. For example, Pronto processes statements like "Bird is a subclass-of FlyingObject with a probability greater than 90%". Pronto is currently a prototype and not ready for prime time, but it could be explored as part of future work.

The main problem is that reasoning is done on already asserted probabilities. There is no function in Protégé that allows counting the number of types of houses and producing their probability values in each urban block individual. Consequently, the uncertainty is calculated outside Protégé and then asserted as the property 'percentage' to link to the relevant data type value. By expressing restrictions based on these percentages, the reasoner can evaluate the membership of the block individuals. We therefore overcome existing DL limitations by asserting the necessary knowledge for the reasoning in this step, which is required to get to our ultimate goal of deriving the high-level concept residential area.

The results of the classification are shown in figures 68 to 70. The dataset with the partitioned building blocks was used to create a thematic map of the classification results. The colour scheme of the blocks is the same as for the underlying types of housing classified earlier. Blocks without buildings are empty and are not useful to the inference procedure. The classification was based on the percentage restriction that an individual block must contain 70% of one type of housing to be classified as an instance of the relevant class. However, the partitioning of blocks is not always ideal. Some partitions enclose quite large areas containing a mixture of differently sized and shaped building features. The results could be improved by using methods that make finer distinctions within the partitions. For example, clustering methods assign objects into

groups (i.e. clusters) so that objects from the same cluster are more similar to each other than objects from different clusters (see chapter 2). This technique could be used to divide blocks that are not homogeneous, thus providing a more fine-grained classification.

Giffnock buildings classified by type of dwelling

	DetachedHouse	(940)
■	EndTerracedHouse	(333)
■	MidTerracedHouse	(944)
■	SemiDetachedHouse	(1769)

Giffnock partitions classified by type of urban block

	BlockDetachedHouses	(18)
□	BlockMixedHouses	(33)
□	BlockSemiDetachedHouses	(40)
□	BlockTerracedHouses	(36)
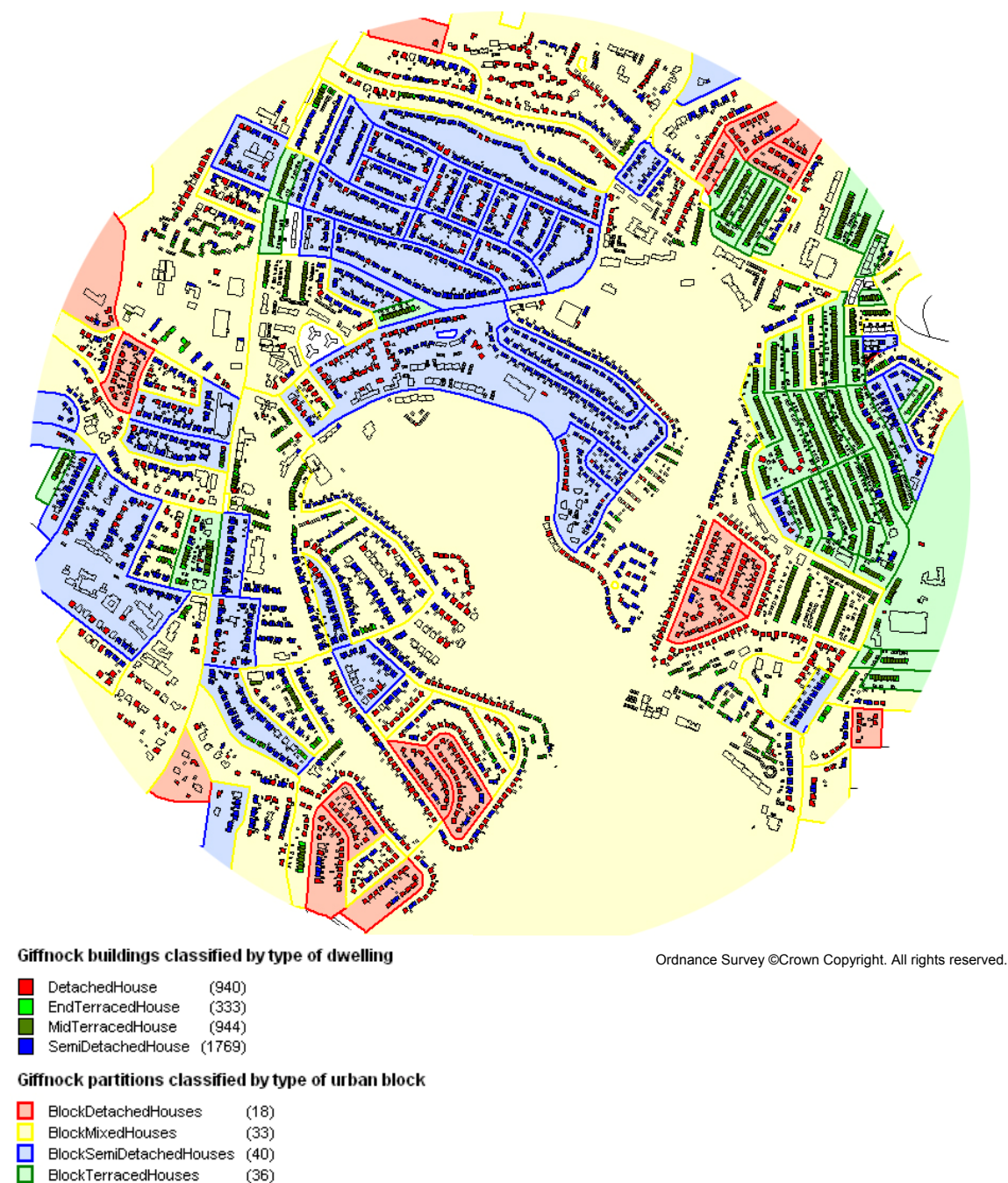
Figure 68 Giffnock inferred types of urban blocks

Drumchapel buildings classified by type of dwelling

- DetachedHouse (400)
- EndTerracedHouse (383)
- MidTerracedHouse (692)
- SemiDetachedHouse (2615)

Drumchapel partitions classified by type of urban block

- BlockDetachedHouses (4)
- BlockMixedHouses (16)
- BlockSemiDetachedHouses (63)
- BlockTerracedHouses (17)
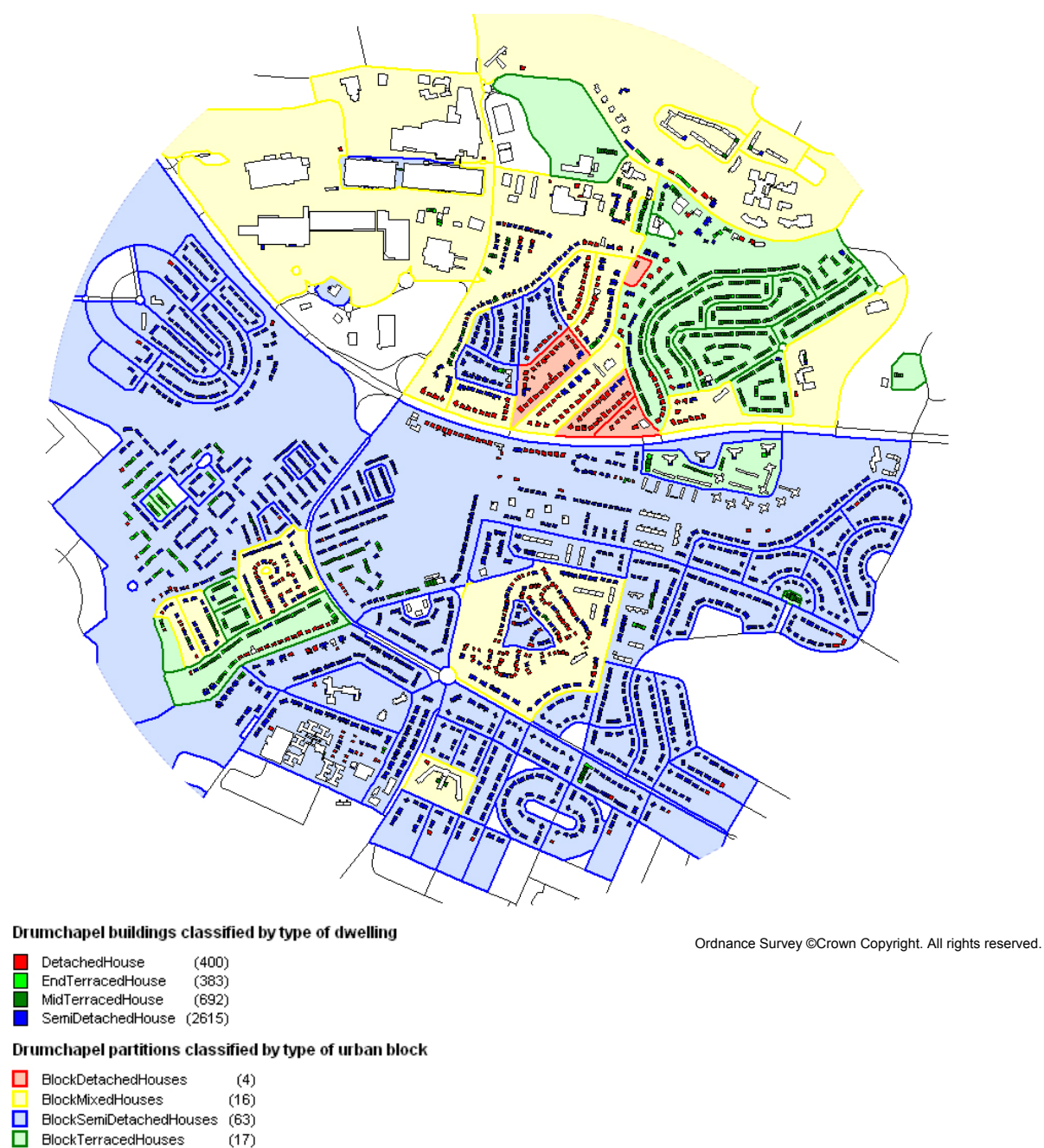
Figure 69 Drumchapel inferred types of urban blocks

Pollokshields buildings classified by type of dwelling
Size threshold set to 280 square metres

- ▮ DetachedHouse (951)
- ▮ EndTerracedHouse (252)
- ▮ MidTerracedHouse (784)
- ▮ SemiDetachedHouse (664)

Pollokshields partitions classified by type of urban block

- ☐ BlockDetachedHouses (24)
- ☐ BlockMixedHouses (47)
- ☐ BlockSemiDetachedHouses (10)
- ☐ BlockTerracedHouses (50)
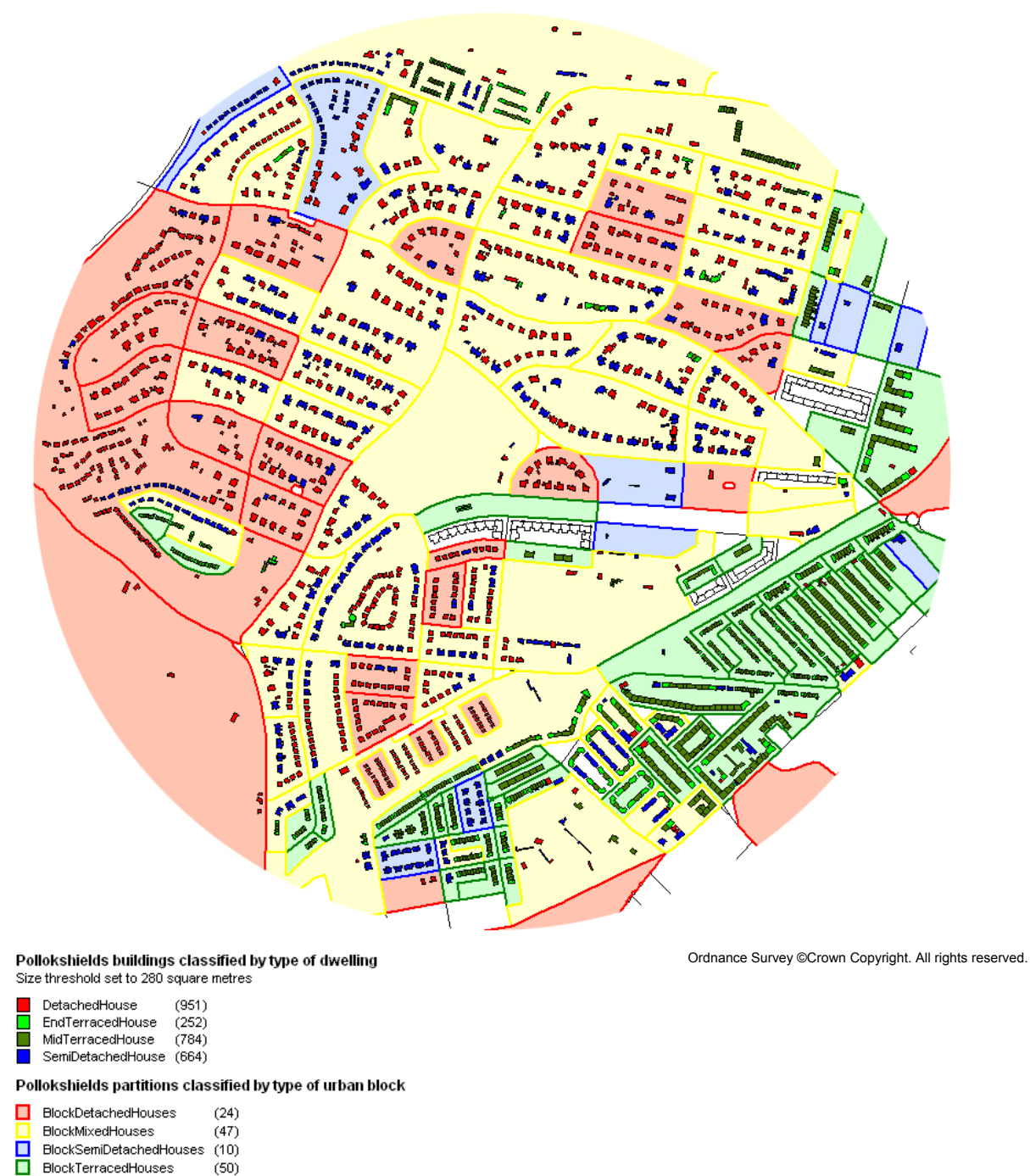
Figure 70 Pollokshields inferred types of urban blocks

Inference of type of district

Districts are visually and functionally characterised blocks (Boffet, 2000). In reality, districts are difficult to define because their limits often superimpose one another, due to their natural fuzziness. However, here we are dealing with a simplified representation of concrete individuals that have been partitioned into blocks. The partitioning of blocks provides the means to define even larger areas of homogeneity such as specific districts. Looking at previous figures with the block classifications, we can already discern districts that are forming out of adjacent blocks that contain the same kind of housing. Interpretation is largely guided by identifying homogenous areas based on Gestalt principles. In this case, similarity and proximity do not relate to individual buildings but to individual blocks. Similarity in this sense is expressed through blocks of the same kind, and proximity is expressed through blocks that are near to one another. By asserting the knowledge of which blocks are connected to one another, we can use the classified types of blocks to define districts. The idea is to assert that if an individual block *a* is connected to block *b*, and *a* is of the same type as *b*, then *a* and *b* are part of the same district. Figure 71 illustrates this idea.
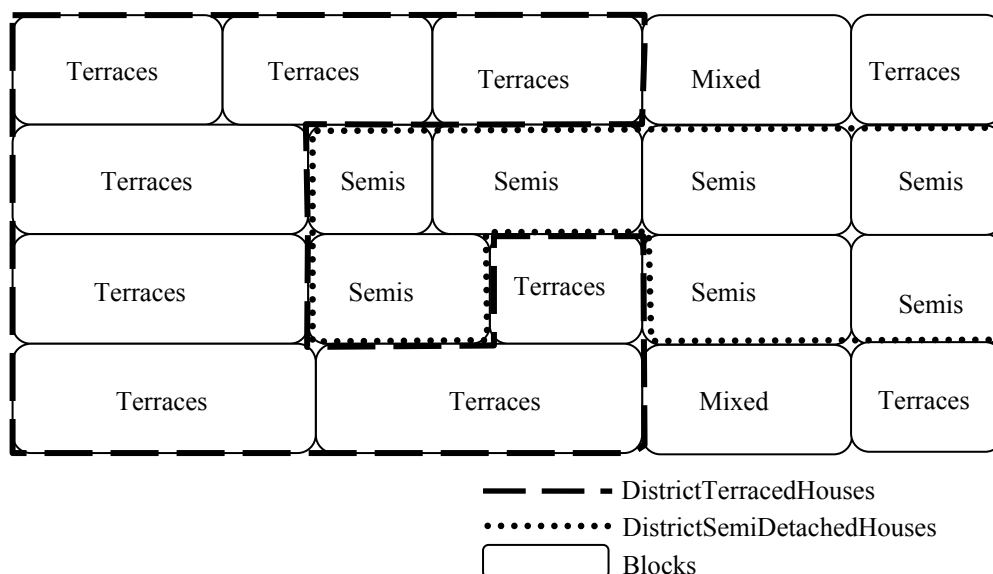
Figure 71 Deriving districts based on connected blocks of the same kind

The reasoner can accurately infer the blocks of the same kind as part of DistrictSemiDetachedHouses despite their connections to other types of blocks. It is important that the property 'connectedTo' is symmetric and not transitive. With a transitive property, a block that is not connected directly to another block of the same kind would still be classified as part of the district, because it is connected via the other

type of block. For example, the two separate blocks of terraced houses in figure 71 would be also classified as part of the class DistrictTerracedHouses because they are linked to the other terraced blocks via the semis and mixed types of blocks. This means if the block 'terraces' is connected to the block 'semis', and the block 'semis' is connected to another block 'terraces', then the first block 'terraces' is also connected to that block 'terraces'. Instead, we want a district defined as a group of blocks with a minimum of two blocks in one district. This issue is eliminated by setting the 'connectedTo' property to symmetric, where the blocks have to directly touch one another.

Figure 72 shows our ontology with the defined district classes. Again, the classified block instances of the previous inference have been asserted as individuals in the ABox. Hence, we can lose the previous class definitions of the types of blocks and keep them as primitive classes denoting the different sets of block individuals respectively. The resulting DL expressivity is $\mathcal{SOIF(D)}$. The way the district classes haven been defined leads to a subsumption relationship to the block classes. For example, the class DistrictDetachedHouses is a BlockDetachedHouses that is connected to some other BlockDetachedHouses. This definition specifies that an individual block becomes a member of the class DistrictDetachedHouses, if it is a member of the BlockDetachedHouses and is connected to another member of the same class. This means all the individuals are members of both the district and the block classes. Defining the districts along the specialisation hierarchy is a simple solution to classifying the block individuals into the relevant types of districts. However, we might want to express that a block is part of district. We could argue that a district of type detached houses should contain 90% of blocks of the same type. In this case the is-a relation is not valid anymore.
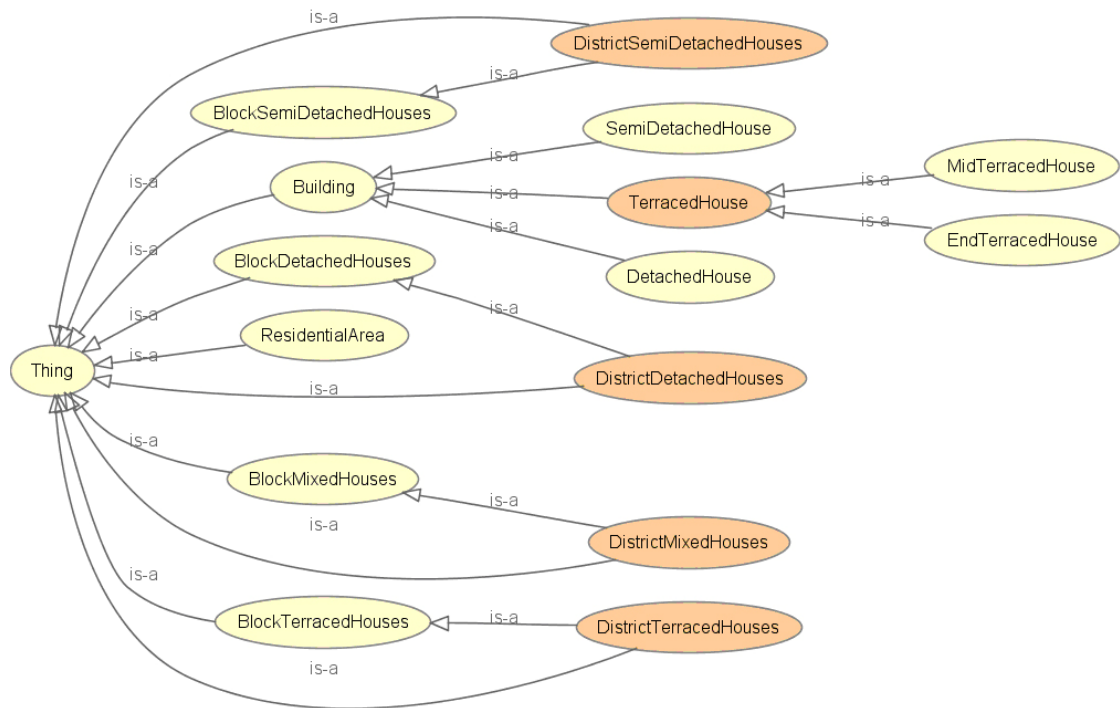
Figure 72 Taxonomy of defined district classes

The results of the inference show why a fuzzy definition could be more desirable. Figure 73 shows an enlarged subset of the Giffnock classification. The classification is overlaid on the previously classified types of blocks and buildings. The districts have a patterned filling to differentiate between the blocks that have not been classified as part of a district. We can see why it might be of interest to change the definition of a district class to include blocks of other types of housing. The classified districts contain only blocks of the same kind. Some individual blocks remain because they do not form a group of blocks of the same kind, such as the single block of terraced houses surrounded by the district of semi-detached houses. In such cases, it might make sense to aggregate a small percentage of different types of blocks into the district, if it forms a uniform whole with the rest of the group.
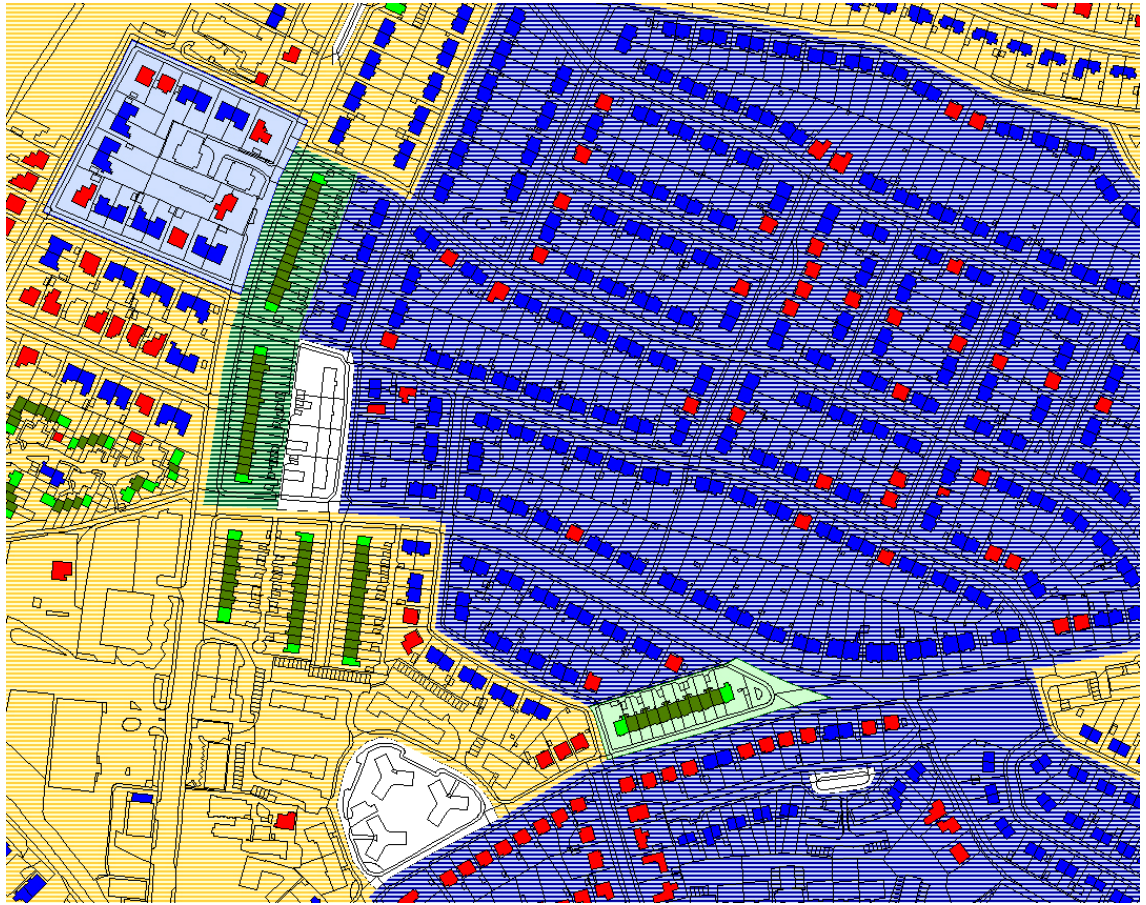
Figure 73 Example of residential districts and blocks (Giffnock)

In conformance with the given definitions of our ontology, the reasoner accurately inferred all types of districts. As we can see in figure 73, the separately located block of semi-detached houses in the top left corner was not classified as part of the class DistrictSemiDetachedHouses. Figures 74 to 76 illustrate the classified districts for all three case studies. Since the inference is based on block partitions, we will find large areas classified as districts of mixed housing. Having more finely delimitated blocks would certainly improve the results at this level as well. Another reason is that the partitioning and inference is based on building features alone. Taking into account other topographic features, such as green spaces, would improve the results overall. However, considering the limited scope of this thesis, these examples illustrate adequately the inference mechanisms at the different abstraction levels of the proposed conceptual model.
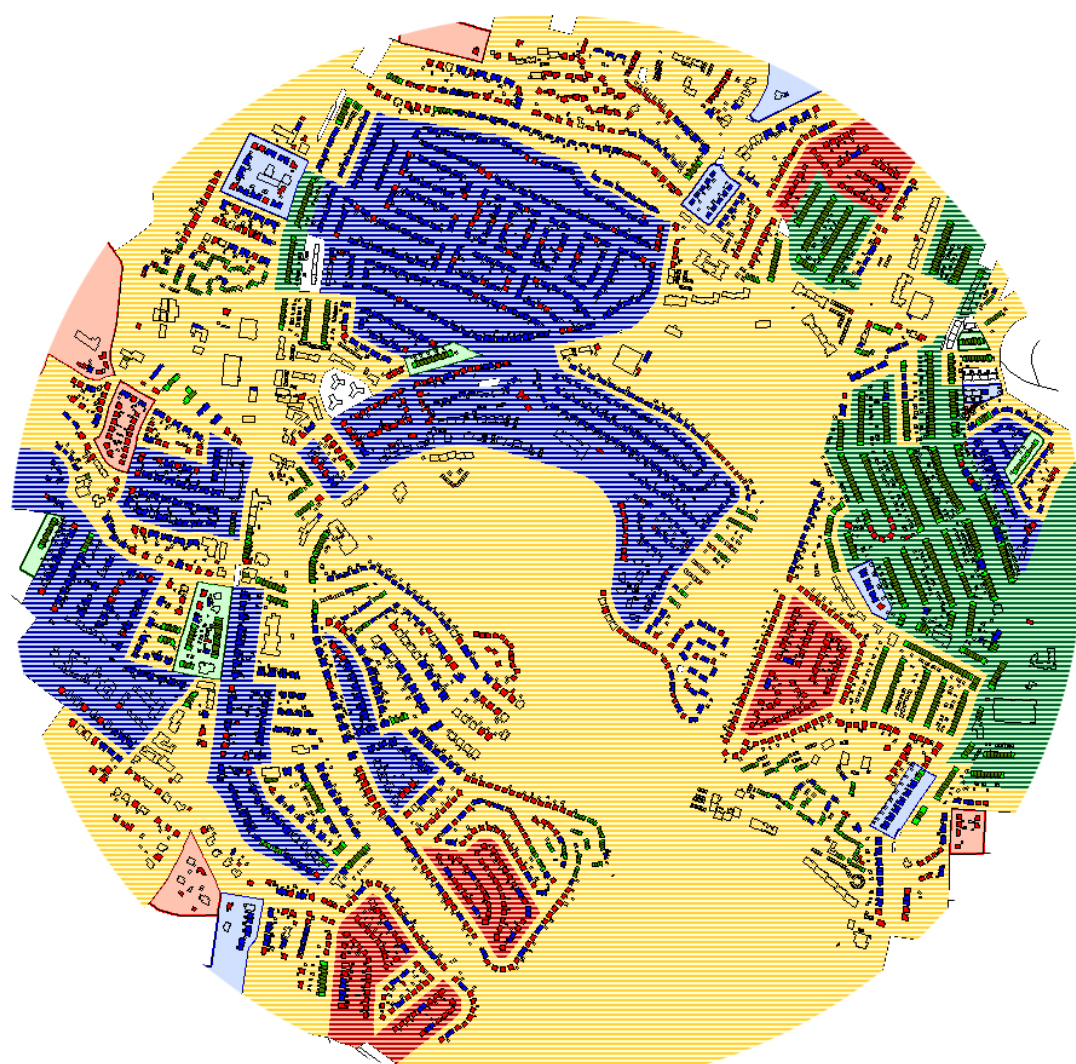
Giffnock buildings classified by type of dwelling

- ■ DetachedHouse (940)
- ■ EndTerracedHouse (333)
- ■ MidTerracedHouse (944)
- ■ SemiDetachedHouse (1769)

Giffnock partitions classified by type of urban block

- □ BlockDetachedHouses (18)
- □ BlockMixedHouses (33)
- □ BlockSemiDetachedHouses (40)
- □ BlockTerracedHouses (36)

Giffnock partitions classified by type of district

- ≣ DistrictDetachedHouses (12)
- ≣ DistrictMixedHouses (33)
- ≣ DistrictSemiDetachedHouses (33)
- ≣ DistrictTerracedHouses (31)

Figure 74 Giffnock inferred types of districts

Drumchapel buildings classified by type of dwelling

- DetachedHouse (400)
- EndTerracedHouse (383)
- MidTerracedHouse (692)
- SemiDetachedHouse (2615)

Drumchapel partitions classified by type of urban block

- BlockDetachedHouses (4)
- BlockMixedHouses (16)
- BlockSemiDetachedHouses (63)
- BlockTerracedHouses (17)

Drumchapel partitions classified by type of district

- DistrictDetachedHouses (2)
- DistrictMixedHouses (12)
- DistrictSemiDetachedHouses (60)
- DistrictTerracedHouses (13)

Figure 75 Drumchapel inferred types of districts

Pollokshields buildings classified by type of dwelling
Size threshold set to 280 square metres

- DetachedHouse (951)
- EndTerracedHouse (252)
- MidTerracedHouse (784)
- SemiDetachedHouse (664)

Pollokshields partitions classified by type of urban block

- BlockDetachedHouses (24)
- BlockMixedHouses (47)
- BlockSemiDetachedHouses (10)
- BlockTerracedHouses (50)

Pollokshields partitions classified by type of district

- DistrictDetachedHouses (17)
- DistrictMixedHouses (42)
- DistrictSemiDetachedHouses (2)
- DistrictTerracedHouses (47)

Figure 76 Pollokshields inferred types of districts

Inference of the residential area

Residential area is a matter of defining a class with a covering axiom that joins the different types of districts. In this case, we can express an axiom where the class ResidentialArea is covered by all types of districts. In Protégé, a covering axiom manifests itself as a class that is covered by a union of classes. Therefore, a covering axiom consists of two parts: The class that is being covered, and the classes that form the covering. Consequently, we can specify that class ResidentialArea is covered by the classes DistrictDetachedHouses and DistrictTerracedHouses and so forth (figure 77). This means that a member of class ResidentialArea must be a member of DistrictDetachedHouses and/or DistrictTerracedHouses. If classes DistrictDetachedHouses and DistrictTerracedHouses are disjoint then a member of class ResidentialArea must be a member of either class DistrictDetachedHouses or class DistrictTerracedHouses. Without a covering axiom an individual may be a member of the class ResidentialArea and still not be a member of DistrictDetachedHouses, DistrictTerracedHouses, etc.
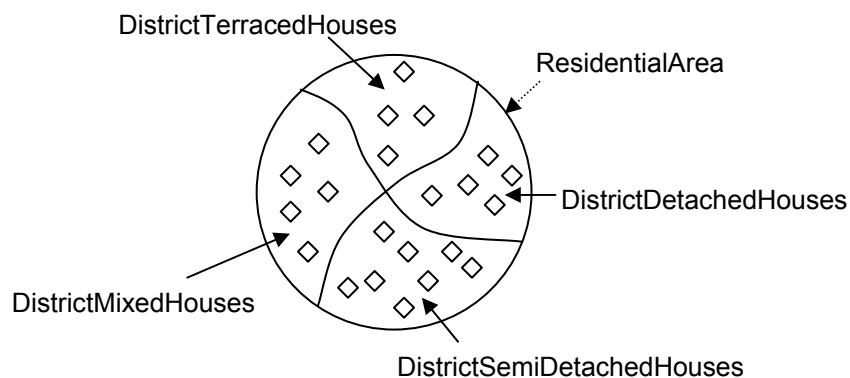
Figure 77 The effect of using a covering axiom

The covering axiom can be extended to include other functions such as recreational areas, i.e., a park. Furthermore, it is possible to refine and split the class ResidentialArea into, say, residential suburbs, inner city or rural areas. These classes could be differentiated by including knowledge about densities, type of housing and relations to other functions such as parks or commercial areas. Cities articulate as spatial patterns with flats near the centre, terraced inner suburbs, and detached/ semi-detached outer suburbs. Often density and distance variables are indirectly reflected in house types (Batty and Longley, 1994). The advantage is that the model can be easily extended to facilitate for such additional knowledge.

As already mentioned, this proof of concept does not include any of the other topographic features. I concentrated on making inferences about the types of buildings and their aggregated concepts. Hence, by applying the covering axiom for residential area, the whole of the sample datasets would be classified as residential area. It requires further work to apply the conceptual framework to other types of land uses, which would then allow us to differentiate between residential areas and parks, for instance, in the given sample datasets. For example, figure 78 shows the park in the Pollokshields dataset. We can create rules that define the characteristics and configuration of a park. From OS MasterMap Topography Layer we know that typical features have descriptive group attributes such as general surface, natural environment, road or track, or inland water. The make is natural and the descriptive term gives attributes such as scrub, coniferous trees and nonconiferous trees. This information can be asserted as part of the ABox individuals including information on size, shape and which features are touching one another. From the questionnaire, we know that people associate pathways, playing fields, trees, hedges and lakes or ponds with the recreational land use (see table 4). We can therefore define a park has having large, unevenly shaped natural surfaces, adjacent to tree areas (coniferous or nonconiferous) and containing paths and potentially inland water features.
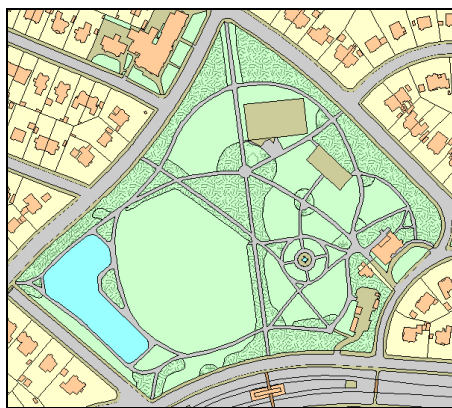
Ordnance Survey
©Crown Copyright. All rights reserved.

Figure 78 Extracting recreational land use example from Pollokshields

Similarly, we can define in detail the industrial area from the Drumchapel dataset (figure 79). Typical characteristics consist of very large buildings (e.g. 7,000 to 12,000 square metres in size), very large man-made general surfaces (e.g. 14,000 to 29,000 square metres in size) adjacent to some natural general surfaces, roads, and railway. An

industrial area requires good access methods for transporting goods, which means such areas are usually located near a railway. It would be worthwhile to evaluate the topological significance of the transport land use, as it is contiguous and connects all other land uses (Marshall, 2005). Indeed, it will be challenging to define rules that make appropriate distinctions between industrial, commercial and even educational land uses because of similar configurations. A detailed analysis is necessary to identify typical features and their characteristics that build up the higher-level categories. Context information provides the key for the differentiation. Whereas an industrial area is likely to be next to a railway, an educational area will contain other features such as a large sports ground. Therefore, to prove the wider applicability of this ontology-based approach to other, more difficult types of land uses requirs further work. Nevertheless, the examples in this thesis illustrate successfully the inference procedures for the residential land use type, where we start out with the database knowledge and a large set of individual topographic features that become classified into higher-level objects.

Ordnance Survey
©Crown Copyright. All rights reserved.
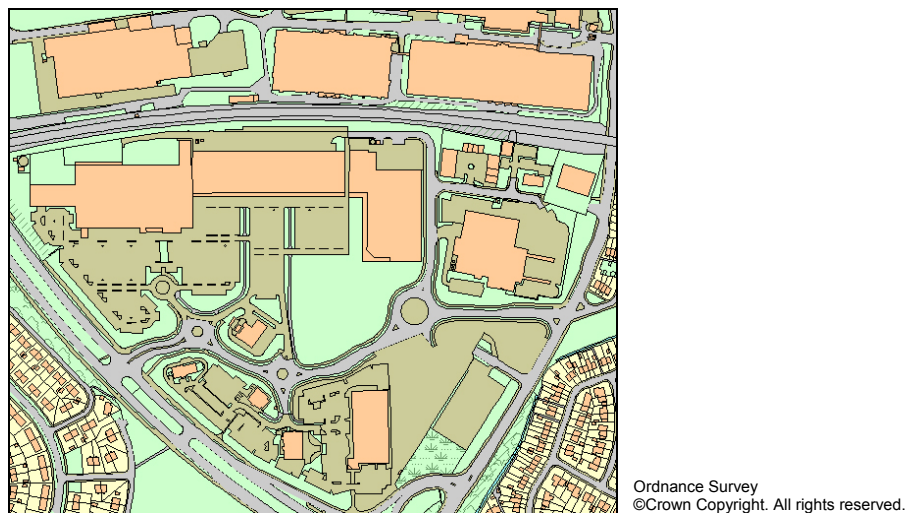
Figure 79 Extracting industrial land use example from Drumchapel

In summary, the procedure of incrementally classifying all individuals into higher-level aggregate concepts resembles the structure of a pyramid (figure 80). At the outset, we have a large store of individuals that were exported from the topographic database along with their attributes and calculated missing relations. With the provided class definitions of the conceptual framework, a terminological reasoner then classifies the individuals by assigning them to their respective member classes. Individuals are then instantiated into increasingly more meaningful, higher-level concepts through a systematic

procedure. The aim is finally to export the high-level annotations such as residential district or area back into the database. As individuals are incrementally instantiated, we are dealing with fewer individuals at the high level (such as block and district individuals). Hence, inference becomes computationally less burdensome the higher up we are the hierarchy.
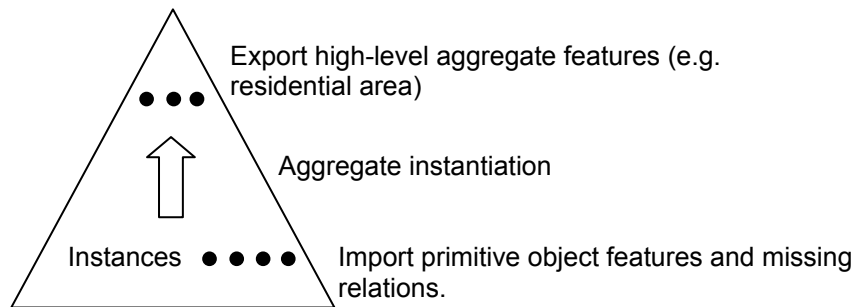
Figure 80 Inferences resemble a pyramid structure

7.3 Discussion

There are clear advantages and disadvantages of using a knowledge-based approach. The following sections discuss the benefits and weaknesses in terms of the applied technologies and how we overcame some of the limitations. An outlook on perspectives of description logic languages places the benefits of this approach into the longer-term view.

Advantages

In general, OWL meets the basic requirements for modelling a geographic domain (Abdelmoty *et al.*, 2005). OWL is a general-purpose language where domains are modelled using user-defined classes and properties. We can therefore represent geographic features and their associated types as well as spatial and non-spatial properties. We can represent the specialisation and generalisation of feature hierarchies and create constraints on the supported types of relationships. Using set operators such as union and intersection, we can define classes through collections of individuals from other classes. In particular, OWL-DL is a good implementable and expressive language, which has received much research over the past years as part of the Semantic Web vision. Formal knowledge representation in general has indisputable merits for analysing the formal structure of a problem and its solution as well as to represent knowledge in a formally accountable way (Neumann and Schröder, 1996).

Knowledge representation provides the basis for powerful inference mechanisms. Protégé for example, provides standard inference services for building and managing ontologies at author time. It offers useful services at delivery time, and it acts as a reasoner at application time. In contrast to information stored in a database, ontologies are much easier to manage as they grow in size because of the subsumption and consistency checking provided by the inference engine. Inconsistencies and conflicting information are detectable and easily tracked down. Despite some deficiencies in query answering, ontologies have the general advantage over standard database query languages (e.g. SQL) that they can infer new information. Their reasoning services can help in the selection of the sources that are relevant for a query of interest. In addition, it can be easier to construct queries over ontologies because the concepts reside on a high conceptual level close to a user's language and understanding.

Fundamental to our problem is that DL classifiers allow composition and instance retrieval. Composition means we can define new concepts systematically from existing concepts. This allows the construction of complex, higher-level concepts out of simpler ones when ascending the interpretation hierarchy of our conceptual model. Since ontology classifiers are designed to reason about the things that are necessarily true about all instances of given types in our conceptualisation (Rector, 2004), we can exploit reasoning tasks for classifying our topographic individuals. In particular, instance retrieval determines the most specific superordinate concepts of a knowledge base for an unknown individual described by attributes and relations. Therefore, if a semi-detached house in a topographic scene is conceptually defined as a 'house that touches max one other house', a classifier can determine whether some topographic evidence satisfies this conceptual description. As we can see from the previous sections, the application of ontologies seems to offer optimistic results for the recognition of new concepts in topographic data.

Consequently, the main benefits include the explicit and more intuitive nature of the modelling. The intended meaning of terms is explicitly defined and expressed through the semantics of the language. Standard reasoning services provide the necessary methods for making inferences about the asserted knowledge. The reasoner gives explanations about its inferences and automatically identifies any conceptual inconsistencies in the knowledge base. With other techniques, such as graph-based or

pattern recognition algorithms, the processes for computing a solution are often hidden in some programming language. In addition, the structure of the model is often obtuse, for example where the representation has been reduced to a number of connected lines (e.g. Hillier, 1996; Béra and Claramunt, 2004). Therefore, alternative methods often suffer from a lack of definition and semantics, as well as the flexibility to make changes to the underlying model at application time.

Lüscher *et al.* (2007 and 2008) try to overcome the limitations of procedural algorithms by combining them with ontologies. They pursue the same goal of accessing higher order semantic concepts such as dwelling types from topographic data. Lüscher exploits ontologies for explicitly describing the properties of, say, terraced house, to inform the recognition process. However, instead of using the reasoning powers of logic-based ontologies for inferring new knowledge, he attaches a piece of code to be executed for computing terraced houses from the vector data. Advantages include that the algorithm can be tweaked to deal with uncertainty, it can learn and tune thresholds on the fly (machine learning), it does not rely on complete information, and it is computationally efficient. Nevertheless, the incorporation of algorithms in such a way does not lend itself very well to the integration in a system with a formal semantics.

Limitations

Unfortunately, a logic-based approach is inherently limited because of its purely deductive nature. Knowledge representation formalisms generally live in a separate paradigm to databases and thus fail to provide the necessary means to directly process database instances. Instead, database properties need to be translated into the abstract logical level that the knowledge representation level resides on. We achieved this through various procedures, by preparing the data in the database, running a python script to populate the RDF syntax with the individuals, and incorporating this syntax into the OWL file. With all these intermediate processes, one has to be careful not to miss any assertions and links between individuals. Errors can creep in quickly, leading to an inconsistent ABox. Although Protégé offers consistency checking, it currently lacks any implemented explanation about individuals, which makes it difficult to find a missing link between thousands of individual assertions. Similarly, we have to transfer the inferred knowledge back into the database. For this purpose, a text parser was used to extract the inferred assertions from the OWL file into a CSV file, which was then

imported back into the database. The whole process is therefore cumbersome and error-prone.

Knowledge representation needs a sound formal basis when the body of knowledge becomes large and diverse. Still, the classification suffered problems when dealing with large numbers of individuals and complex role assertions. A geographical application domain can potentially contain hundreds of concepts and many thousands of individuals (Abdelmoty *et al.*, 2005). For example, the reasoning for this proof of concept was carried out on a 2GB RAM, 4CPU, 3GHz workstation. If the ontology became too complex in terms of a large ABox and complicated expressions, the inference process was either very slow, taking anything up to a couple of hours, or it ran out of memory completely. If sufficient memory was available, then the processing would only be a matter of a few minutes. Therefore, a major limitation is not only the failure to access the database layer directly, but also the inability to query over large sets of individuals. This, however, is a general problem. Large datasets of greater structural complexity usually lead to computational inefficiency and in some cases to greater uncertainty (e.g. Barnsely et al., 2001; Conroy Dalton and Kirsan, 2005).

Precautions can be taken to avoid the computational complexity of large ABoxes and rich concept definitions. Firstly, the number of individuals was reduced by excluding building features of size less than $35m^2$. Secondly, the concept definitions were simplified. In some cases of the classification, this meant there was insufficient knowledge to correctly infer all types of dwellings. It therefore remains a diligent trade-off between expressivity and deductive power of the DL system.

Another challenge is the modelling of concrete domains. Currently, we cannot sufficiently handle spatial aspects like topological and distance relations with OWL-DL, except for linking properties to data types such as float, string, integer, etc. As spatial data are a concrete domain, it is yet not possible to directly infer knowledge from spatial features. Tools providing such inference services are not yet available and further research is still required to this end. Hence, qualitative relations needed for the conceptual modelling must be instantiated outside the DL system because almost all terminological systems have no built-in primitives to support spatial or temporal reasoning. For example, there is no efficient access to spatially adjacent objects unless

one provides user-defined generator functions (Neumann and Schröder, 1996). Consequently, we had to adopt a custom approach to calculate the necessary missing information on spatial relations and block individuals outside the DL system. Although we can criticise such procedural attachments to ontologies (e.g. Lüscher *et al.* 2008), it seems that currently it is not possible to achieve a complete high-level interpretation within a formal DL system.

The DL paradigm imposes many restrictions on the modelling of our domain. The asserted individuals and the domain must be closed to enable reasoning, which goes against the open world assumption of DLs. This leads to insufficient query processing. Another aspect is that the inference procedure advances along the specialisation hierarchy, whereas an aggregate should ideally be formed from parts as described by the conceptual framework. However, the formal semantics of parts and wholes is problematic (Neumann and Schröder, 1996). It is difficult to express that parts become something special when they constitute an aggregate. In other words, one of the problems is to induce a classifier to assemble suitable parts into an aggregate. Description logics do not offer a pre-defined part-of role like many frame systems, and therefore it would be interesting to implement the conceptual model in a frame-based ontology (Wang *et al.*, 2006).

A further, yet more general problem is the modelling of fuzziness. Regions considered in geography often do not have crisp and well-defined boundaries. Whether we model spatial regions in a GIS using the region connection calculus (RCC), or whether we conceptually model regions through qualitative spatial relations in a knowledge base, we are still abstracting away from the complicating aspects of reality. Although progress is being made in terms of reasoning over fuzzy concepts, e.g. Pronto (Klinov, 2008), OWL does not express fuzzy or vague concepts (Goodwin, 2005). Presently, there is no hypothesis generation, no guessing of likely classifications, not even a computation of possible classifications. There is no mechanism to compute missing evidence as it is needed, such as in calculating the percentage of types of dwellings in each asserted block individual. Classifications are deduced from evidence that must be completely provided beforehand.

Overall, knowledge engineering is hard. There is rarely one person who is both a domain expert and ontology expert (Goodwin, 2005). Those who have not yet used a terminological system will probably need time to get used to the logical expressions. Knowledge acquisition often forms a bottleneck in the progress of knowledge-based techniques (Weibel *et al.*, 1995). Although we are trying to generate a solution that reflects the human conceptual nature and cognitive perception of patterns, we still rely on prior information, thresholds and parameters. Hence, similar to other classification methods (e.g. Boffet, 2000), the chosen thresholds have to be sufficiently sensitive to discriminate significant classes. For example, a small threshold value for the class House results in many omissions. On the other hand, a large threshold value results in many misclassifications. Improvements can be achieved by combining and augmenting data sources, knowledge and methods. We can assert additional knowledge to refine the class definitions. Alternatively, we can include other methods, such as clustering techniques (e.g. Anders *et al.*, 1999), to get a more fine-grained classification of the types of blocks and districts.

Perspectives

Description logics are versatile as they play a key role in may applications ranging from medicine, databases, semantic web, to geographic information science. The increasing use of DL based ontologies already stretches the capabilities of DL systems in terms of modelling quality and performance, and thus brings with it a range of challenges for future research. In response to users' requests, DLs are continuously being researched, improved, and extended. There will be increased expressive power, improved scalability, extended range of reasoning services (e.g. explanation, matching, approximation), and hybrid systems are being developed for reasoning more efficiently over spatial data (Cuenca Grau *et al.*, 2006; Wessel and Möller, 2007; Grütter and Bauer-Messmer, 2007a). The tools and infrastructure are also continuously expanding with open source communities such as Protégé that will deliver support for large scale ontological engineering and deployment in the future.

Only recently, a new W3C Working Group formed to work on the next OWL language, which came into life as OWL2 in April 2008 (Cuenca Grau *et al.*, 2008). The new design of this language increases language expressivity (compatible with the description

logic $\mathcal{SROIQ}$), adds property and qualified cardinality constructors, extends data type support and annotations, and includes simple meta modelling (Motik *et al.*, 2008). Similar to the earlier version of OWL, OWL2 has profiles that place restrictions on the structure of OWL2 ontologies. They are trimmed down versions of OWL2, which trade some expressive power for the efficiency of reasoning. For example, OWL2-QL is aimed at applications that use very large volumes of instance data, and where query answering is the most important reasoning task. These languages will be the future for query optimisation as well as new database developments such as Oracle, which is starting to incorporate to some extent the RDF and OWL data model (e.g. Lopez and Annamalai, 2006).

In addition, there is a growing body of research about spatial knowledge and related reasoning services, which will potentially overcome current limitations in spatial reasoning (e.g. Grütter *et al.*, 2008; Grütter and Bauer-Messmer, 2007b; Katz and Cuenca Grau, 2005). It is possible to envisage the development of specialised tools for manipulating elements in the ontology. We therefore have to see the value of this approach in the longer-term according to advances in artificial intelligence and inference mechanisms. In comparison, Lüscher's (Lüscher *et al.*, 2008) approach of using ontologies to describe pattern recognition algorithms, for example, is shorter term since it relies on algorithms that operate directly on the data. Although we also had to rely on calculations made outside the DL system to support our inferences, most of the encountered limitations are of technological nature. It can be trusted that in the near future most of the difficulties and incompatibilities identified throughout this thesis would be overridden by the evolution of systems and the refinement and enrichment of ontology languages.

Conclusions

The underlying question of this thesis asks what types of functional information can be derived from topographic data alone. The thesis pursued a method that starts with the land cover parcels and spatial structures stored in a database to incrementally reason about higher-level land use information. According to Clawson and Stewart (1965) characteristics of a good land use classification are a pure line classification that describes activities only, a system that is useable in detail as well as in summary form if

desired, and a classification that is based upon what you actually see on the ground or on the map. The conceptual model that this thesis proposes meets all of these criteria: The model is pure because it models only relevant concepts. Flexibility is provided through the different levels of granularity that the model represents. The model builds upon people's conceptualisation of the land use domain, which is reflected in its high-level concepts. Because the model is implemented with a knowledge representation formalism based on logic, it is readily susceptible to machine processing and we can apply standard inference services. Lastly, because the model's explicitly asserted semantics are dissociated from the database layer, it promotes interoperability and transparency.

The characterisation of land uses and urban patterns enriches the topographic data and helps to improve the map generalisation of buildings. However, most of these higher-level characterisations are not specific to a generalisation purpose, but can be used for other applications such as in urban studies (Gaffuri and Trévisan, 2004). The results presented in this chapter are specific to residential land use, and illustrate how semantic reasoning can be applied to topographic data to semantically enrich its thematic contents. Although it needs to be proven, the proposed conceptual framework is potentially valid for other types of land use. The exact formulae and properties may have to be modified and additional knowledge has to be asserted, but the underlying framework remains the same. However, creating the mappings between the data layer, the real world, and a knowledge representation language is a labour-intensive and error-prone activity. Many current mapping tools are semi-automated, helping humans in an interactive manner. In particular, the mapping between semantically lightweight representations (e.g. spatial data) versus semantically rich representations with formal axiomatisations (e.g. OWL) still requires a trade off between computational cost, flexibility and powerful reasoning capabilities (Uschold and Grüninger, 2004).

It is important that the conceptual model is consistent with the phenomenon under investigation, that is, it must aid and not hinder the explanation of the phenomenon under investigation. With description logics, we have the advantage of precise conceptual definitions with well-defined semantics. The creation of high-level structures always requires abstraction, and such abstraction should provide a set of guiding principles, which select, organise and order relevant elements, independent of

contingent factors. With description logics, we can specify a method of analysis and define the relevant variables. A DL system's reasoner provides a standardised way of processing knowledge and making deductive inferences asserting new facts. It is therefore easier to maintain the classification rules, and it provides flexibility for modelling differences between regions as well as possibilities to adapt to future requirements (Hartog *et al.*, 1999). For example, you can define new terms for special uses based on the existing vocabulary, in a way that does not require the revision of the existing definitions. The conceptual framework therefore should offer extendibility so that one can extend and specialise the ontology monotonically (Gruber, 1993).

The main advantage of knowledge engineering over programming is that it requires less commitment, and thus less work (Russell and Norvig, 1995). A knowledge engineer only has to decide what objects and relations are worth representing, and which relations hold among which objects. A programmer has to do all that, and in addition must decide how to compute the relations between objects, given some initial input. The knowledge engineer specifies what is true. The inference procedure then figures out how to turn the facts into a solution to the problem. Furthermore, because a fact is true regardless of what task one is trying to solve, knowledge bases can, in principle, be reused for a variety of different tasks without modification. Hence, in view of the complexity of hand-coded classification processes, it would be an advantage to make use of a classifier offered as an inference service of a terminological system (Neumann and Schröder, 1996). This would not only save software development efforts but the formal semantics of the terminological system would facilitate knowledge reuse through the use of ontologies. Furthermore, ontologies provide a precise account of what we want to model or, as in this case, recognise in the data. Assumptions that would remain implicit in informal definitions have to be spelled out. Nothing is hidden and inaccessible within a knowledge-based system. The model is explicit and can be easily changed or adapted to new application contexts. Indeed, the scope of concept definitions potentially suffers from a lack of expressiveness of terminological languages, as reasoning processes may become computationally more complex or even undecidable. Nevertheless, because of a logic's inference processes, debugging a knowledge base is made easier by the fact that any given sentence is true or false by itself, whereas the correctness of a program statement depends very strongly on its context. Inference

procedures of consistency and satisfiability checking ensure that the ontology is consistent automatically.

There is no doubt that we need reasoning services that are better equipped for interpretation tasks. To make progress, individual researchers and practitioners will have to initially make many assumptions, and then relax them one by one as technology progresses. Some may argue that given the widespread importance of knowledge representation to the field of cognitive engineering, that such efforts are bound to be fruitful, regardless of the findings they produce. According to Horrocks (2005b), the effective use of logic-based ontology languages in applications such as this one will critically depend on the provision of efficient reasoning services to support both ontology design and deployment. This, however, is a technical limitation and not a conceptual one.

Chapter 8

Conclusion

"Masses of low-level data are all very well, say the artificial intelligence researchers, but unless and until these can be presented in a form that is humanly intelligible – making use of the high-level concepts that typify human qualitative spatial reasoning – they can never be put to good use outside a rather narrow range of technically-motivated concerns. Your abstract high-level theorising is all very well, reply the people who work with low-level data, but how can any of it be applied in practice?"

–Antony Galton (1999, p.251)

In the view of the conflicting attitudes between artificial intelligence researchers and those concerned with low-level data, this thesis takes a small step towards combining both directions with the aim to bridge the gap between the higher and lower level approaches to spatial information. Artificial Intelligence is a different way of looking at the world and it requires a willingness to experiment. Perhaps we need divorce ourselves from traditional methods and technologies to do best. The fact is that there is considerable difference between users' interest in reality and the map contents described by using only the low-level perceptive features. The problem that we are faced with is the lack of semantics in spatial databases, and with that inherently the lack of flexibility and interoperability. The obvious solution is to enrich data sources to equip them better for real-world applications. This thesis exploits high-level conceptualisations for this purpose. Land use information is regarded as a high-level concept that is in most cases implicitly represented within the spatial configuration of features stored in a topographic database. Implemented through knowledge representation formalisms, a conceptual model allows reasoning about and assigning semantics to spatial data. The semantic gap is filled by classifying low-level visual features according to the high-level concepts of the model, thus exposing new, previously implicit knowledge within the data.

This chapter summarises the main research findings in the next section. It discusses the methodological, conceptual and technical implications of this thesis, and highlights its

contributions in each area. With a critical view on what this thesis has achieved, I describe potential benefits and its impact both from an applied and theoretical perspective in section 8.2. Although the thesis attempts to answer many questions, it inevitably poses new ones that need to be addressed with future work. Section 8.3 addresses possible research avenues that can be taken from here onwards.

8.1 Summary of the research

Ontology started as a philosophical notion some 2000 years ago when Aristotle first began to analyse syllogisms. Now it is part of artificial intelligence in terms of building cognitive models for automated reasoning. Ontology addresses the high-level conceptualisation of the world, and thus offers promising aspects for modelling and reasoning about high-level functional concepts in regards to low-level spatial data. In particular, the model offers an instance-based approach to generating inferences on discrete spatial information. For example, OS MasterMap provides classification for individual features such as buildings. In principle, it stores land cover information. Although cartographic text exists to identify the location of functional sites such as a school, there is no explicit association between individual features and the complex features. This thesis attempts to make such higher-level, complex functional information explicit based on the example of residential land use. To achieve this, the thesis addressed the following research questions:

1. What can spatial context and its configuration tell us about the functioning of its features?

2. What can we learn from our own abilities to interpret land use information from topographic maps? What kind of knowledge and reasoning processes are required?

3. How can people's knowledge be captured and transformed into machine-readable format?

4. How can we bridge the gap between knowledge, i.e., conceptualisation, and geographic data, i.e., representation?

5. How can geographic space be modelled in terms of its context and arrangement?

6. What types of functional information can be derived from topographic data alone?

Previous research, especially in urban studies, shows that there is a special relationship between spatial form and spatial function. This relationship forms the fundamental hypothesis of this thesis: Functional information is implicitly represented within the spatial configuration of topographic features (chapter 2). The next question is how we can make this type of information explicit within spatial data. The best interpreters are us human beings. Hence, it suggests itself to investigate the way people interpret this information from topographic maps and how they conceptualise the land use domain in relation to its underlying landscape. Chapter 3 provides the results of this investigation, which gives a clear indication of how successful any automated approach could be – with some land uses (e.g. residential) being easier to determine than others (e.g. educational).

Interpretation is a knowledge-intensive task that requires background knowledge and experience to categorise new observation data. Ontology offers a way to capture, model and transform our acquired knowledge into machine-readable format (chapter 4), thus answering the third question. However, we still have to bridge the gap between the interpreted high-level functional concepts and the low-level data. Knowledge about space consists of the recognition and elaboration of the relations among geographic primitives and the advanced concepts derived from these primitives (Golledge, 2002). Therefore, the thesis proposes an agglomerative approach, where higher-level meaning is instantiated by combining individual features into more meaningful objects. Similar to the interpretation process where groups are determined through their similarities and proximities, the model instantiates increasingly more meaningful objects from types of dwellings and urban blocks to residential districts (chapter 5). The recognition of the whole map arises from the recognition of its parts, which are defined by their underlying data structure.

Through an ontology and its knowledge representation language, we can model the hierarchy of different levels of abstraction, the relations between individual objects and the high-level concepts to be made explicit within the data (chapter 6). Hereby, we have to take care of the duality of the problem in terms of the low-level data descriptions and the high-level concept definitions. This knowledge is formally represented in the ABox and TBox of a description logic system, respectively. The TBox consists of logical predicates that allow the composition of further predicates by logical connectives and

quantifiers. These formal structures then receive intentional meanings and are used to classify the asserted individuals of the ABox, which map directly to the topographic features in the database.

The proposed framework is applied to a sample dataset from OS MasterMap topography layer to identify if functional information can be practically inferred from topographic knowledge alone. For this purpose, a proof of concept is implemented in Protégé 4 Alpha using the ontology language OWL-DL (chapter 7). Based on the standard reasoning services of description logics, a reasoner infers which individuals from the ABox are instances of the defined classes in the TBox. The approach successfully infers the instances for the different types of dwellings from which it then infers the instances of types of blocks followed by the instances of types of residential districts. Figure 81 gives an overview of the thesis, which consists of a conceptual part and the system architecture. Next, we look at the main strengths and weaknesses of this approach, which highlight the implications of the methodological, conceptual and technical aspects of this thesis
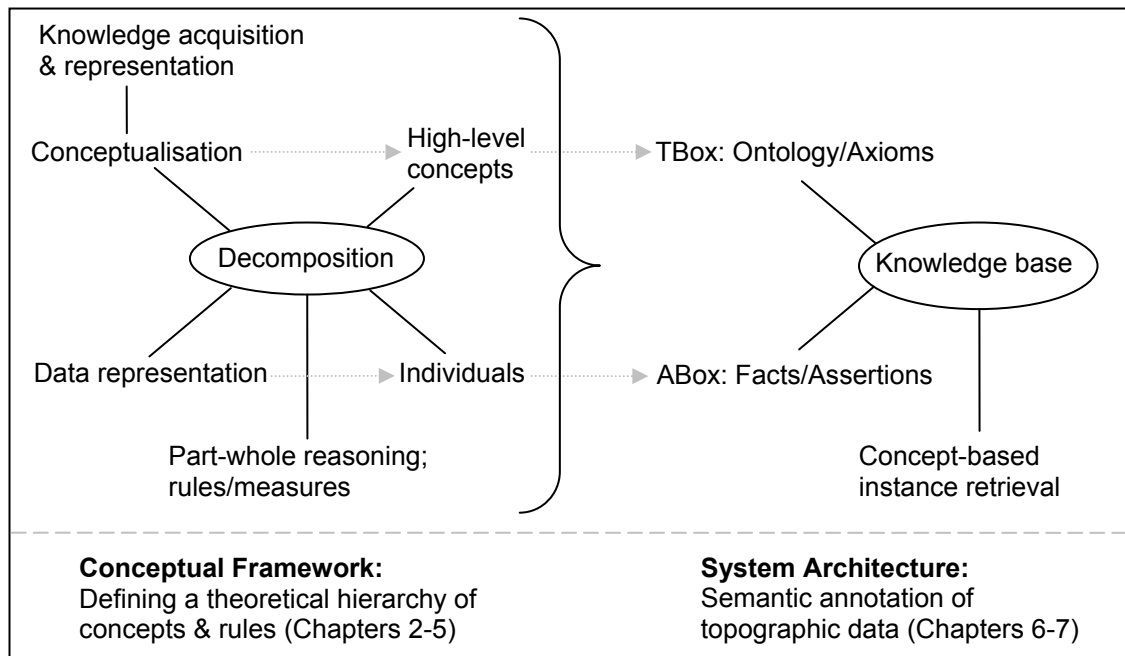
Figure 81 Summary of the thesis

Methodological conclusions

The thesis develops a methodology to infer higher-level information from a topographic database. Interpretation, or inference of higher-order meaning, is a knowledge intensive task, and it has been widely acknowledged that research on information extraction must consider primarily the semantics of the data (chapter 2). The proposed methodology treats inference as configuration problem solving. It conceptualises and uses human knowledge to determine the function of individual topographic features according to their surrounding context. Humans use multiple mental models of the world to reason efficiently at different levels of abstraction (chapter 3). The thesis decomposes this process into its elementary abstraction levels – from individuals and blocks to districts and neighbourhoods – to link between low-level representations and high-level interpretations. This method correlates a one-to-one mapping between rich, semantic knowledge and the syntax of land cover objects, thereby producing a model that is broad enough to capture high-level concepts, but also fine-grained to account for the level of detail given in low-level representations.

GIS research must separate the conceptual database schema from the physical storage arrangement and link it to a third schema describing subsets of the conceptual view according to users and their specific tasks (chapter 4). It therefore seems most promising to combine a top-down approach, from the human elicited conceptualisation of the land use domain, with a bottom-up approach that originates from the representation of topographic data. This middle-out approach incrementally links different levels of details, and thereby derives a coarser description from a more detailed one. The semantics specifies the context for each abstraction level based on a set of relationships that have to be fulfilled by individual features. Ontologies capture and structure this knowledge. By assigning intentional meaning to the concepts that we wish to recognise within the data, ontologies can be used to classify topographic instances into their respective higher-level classes. In this sense, ontologies form the core of the mediation-based approach to information integration, which not only allows the handling of semantically heterogeneous datasets but the inference of implicit knowledge.

The treatment of high-level concepts as neighbourhood structures with their flexibility to form organically proves in many cases to be a better solution to portraying high-level information (e.g. Wahl, 2008). For example, a lot of existing land use data is in

statistical form that confines to a fixed quantity of administrative districts or wards. In this thesis, districts grow out of the characteristics of housing, thus building upon what the data contains. Furthermore, the methodology embraces geographical concepts that are shared in common by non-experts. It brings the advantage that it is more likely to render the results of work in geospatial ontology compatible with the results of ontological investigations of neighbouring domains. It has advantages also in more immediate ways, above all in yielding robust and tractable standardisations of geographical terms and concepts (Mark and Smith, 2001).

The methodological drawback of this approach is the lack of procedure or guideline that can help in the acquisition of knowledge required for a modelling task (Lind, 1999). The elicitation of knowledge from people (experts as well as non-experts) is not easy, and can be subjective and error-prone. Human beings perceive and recognise on a level of near unawareness. This makes it hard to reveal the underlying processes and transfer them to a machine. For example, the questionnaire survey in chapter 3 asks a person to use his or her cognitive mapping skills so that he or she can express to us the characteristics of the very same cognitive mapping process. According to Downs and Stea (1977), this poses serious problems for research into the process of cognitive mapping because the translation into written word is masked by skills to do so. Therefore, the ability to translate knowledge makes it difficult to say this is how a person knows land use.

As a result, building the conceptual model is a time-consuming and versatile process with no singular correct way of doing the modelling. There is no process for model building or for revising, modifying and validating a model. In the worst case, the model may only rely on the modeller's knowledge if not wider knowledge was acquired to ensure the acceptance of concepts across a user community. This potentially leads to a domain bias as well as a mechanism bias, where particular elements are initially selected for examination based merely on the assumptions of the modeller. Currently, these deficiencies are amplified by the circumstance, that the meanings of the different levels of abstraction in the model are only defined in terms of prototypical examples from the topographic domain. This is reflected in the choice of data that the conceptual model builds upon. For example, the thesis focuses on topographic data from Great Britain provided by Ordnance Survey MasterMap. The manual interpretation as well as

the set of automated inferences takes place in the context of a specific location in reality, which is constrained by local surroundings. In other words, the spatial arrangement, composition and context of the geography treated by this thesis may vary greatly to the environments of other countries (Steiniger, 2006). Therefore, the outlined process for model generation is not general and can potentially vary, as conceptualisations may require adjustment in their specification when different datasets are treated. It may not be easily transferred to other problem areas such as for other types of land uses.

Conceptual conclusions

The thesis adapts an empirically grounded theory that starts from data that are broken down, conceptualised, and put back together in new ways to generate a rich, tightly woven, explanatory theory that closely approximates the reality it represents (e.g. Hereth *et al.*, 2000). Land use is a high-level abstract concept, but it is also an observable fact intimately tied to geography. The thesis decomposes this relationship and provides representations of geographical features in the way they are partitioned according to Gestalt principles, that is, in the way people interpret topographic maps for land use information. Because all information ultimately rests on observations, semantics are physically grounded in processes and are mathematically well understood. Exploiting this foundation to understand the semantics of information derived from observations produces powerful semantic models. With such models, we can then reason about the described phenomena and derive new knowledge. Considering the current gap between data representations and high-level conceptualisations, this kind of approach is needed, not only to make sure that representational and modelling languages are compatible but also that models become more intuitive.

Ontology is considered as a strictly pragmatic enterprise. It concerns itself not with the question of ontological realism, that is, with the question whether its conceptualisations are true of some independently existing reality. Rather, it starts with conceptualisations, and goes from there to a description of corresponding domains of objects or closed world data models. This can be interpreted as a failure because ontology is based on a methodology that ignores the real world of flesh-and-blood objects in which we all live, and focuses instead on closed world models (Smith and Mark, 2001). However, this is not necessarily a bad thing. Closed world models are much simpler targets, from a

mathematical point of view, than their real-world counterparts are. In particular, when implementing a formal system for processing database knowledge, we are forced to deal with closed world assumptions (chapter 7). Ontology in this sense provides the best link between closed world models and conceptual representations, which attempt to represent the real world more intuitively. However, some may question the ability of formal ontologies to provide conceptualisations that are more intuitive in the first place.

Ontological engineering is based on model-theoretic semantics (chapter 6). Kuhn (2005), for example, argues that model-theoretic approaches are limited in their meaning because they are restricted to sets. Unstructured sets are too weak to serve as interesting conceptualisations of the world. Especially humans do not understand domains as sets of things and subsets formed by predicates, but through their behaviour and the actions that can be performed in them. A more fundamental pitfall of model theory, in Kuhn's eyes, lies in the symbol grounding problem. Grounding the meaning of symbols through symbols is an oxymoron. Meanings are not fixed and cannot be assigned to symbols independently of how these are used. Therefore, all symbolic approaches to semantics are necessarily limited in scope and need to be complemented by studies of language use and evolutions. It boils down to accounting for meaning by modelling observable effects in the world. This can be achieved by grounding the ontology in the real world and aligning its concepts to people's conceptualisations of the domain (chapter 3).

Without going into further depths of the philosophical issues surrounding ontologies, we have to accept that they simply are another representation, a surrogate, for the real world. All surrogates are imperfect, and from this, two important consequences follow (Davis *et al.*, 1993): Firstly, in describing the natural world, we must omit some of the effectively limitless complexity of the natural world. This means the conceptual descriptions of the model are reduced to what is required for the reasoning. Secondly, if the world model is somehow wrong – and all representations are imperfect – some conclusions will be incorrect, no matter how carefully drawn. Therefore, despite the reasoning services that knowledge representation formalisms offer, drawing only sound inference does not free reasoning from error. It can only ensure that inference is not the source of the error.

Despite these flaws, simplification does not necessarily imply limitation. We have to accept the capacities of any form of representation. Ontologies confine representation to a hierarchical collection of concepts. Hierarchical representations have been long criticised by researchers in terms of their inability to portray a given domain accurately. In particular, the geospatial domain requires modelling that extends beyond hierarchies. Land uses, for example, are composed of overlapping areas that are more lattice- than tree-like (Alexander, 1965). Equally important are therefore non-taxonomic relationships, for example that houses and gardens are parts of residential areas. Reasoning with these is much harder because it is not the simple set inclusion kind required for taxonomies, but depends on the semantics of each relationship. The conceptual model that this thesis proposes recognises this fact, and treats the land use domain as a configurational problem that consists of parts and wholes (chapter 5). However, in its implementation (chapter 7), the model's inference still takes place along the specialisation hierarchy.

The main advantage of the conceptual model is its dissociation from the data. It provides hooks that allow for a direct link, but its conceptual descriptions will not be affected when changes occur to the source data. Equally, the conceptual model offers flexibility because if it misses concepts for describing a specific situation, it can easily introduce new concepts and develop new data components. This is an additive process that would not necessarily alter existing definitions or structures (Hart and Greenwood, 2003). The model builds upon people's conceptualisation of the land use domain, which is reflected in its high-level concepts. Yet, the model's levels of spatial form and spatial function clearly refer to the spatial extension and behavioural characteristics of physical objects (chapter 5). It is unclear whether the definitions of these levels leave room for other types of entities like temporal processes (e.g. land use change). The relation implied by the model between spatial form and function is therefore only valid for material objects found in the geography of the real world. Relations between an action and its attributes are not represented.

Overall, there is the danger of the developed ontology to become another island in a sea of different conceptualisations, which are hard to connect (Kuhn, 2005). We have to take care not to develop ontologies that lack the means to ground conceptualisations in reality. For this reason, it is important to establish a link between the conceptual model,

people's conceptualisations and reality. This requires the study of how people conceptualise the real world, as attempted by the questionnaire survey in chapter 3, and to incorporate these views into our computer models. Otherwise, our models depend on us for their interpretation and we have no account of how meaning gets into the system. The applied evaluation in chapter 7 focused on the inference procedures. However, the categories used in the ontology could be adapted to reflect better the concepts used by ordinary people to describe land use types (e.g. table 3). Without solving this symbol grounding problem, ontologies cannot anchor their conceptualisations in reality and their usefulness remains questionable.

Technical conclusions

Despite lingering controversies among researchers, the thesis develops a method in favour of ontologies. It defies both the attitudes of baseless enthusiasm and deligitimating rejection of ontology, and instead takes a practical approach in terms of what can be realistically achieved with ontologies. Thereby, the thesis goes beyond pure conceptual work (e.g. Mennis *et al.*, 2000; Peuquet, 1988) by considering how objects and classes are actually generated from observational data. With the aim to enrich a topographic database with functional information, the thesis contributes with a systematic approach of converting measurable spatial database properties into high-level semantic information. It successfully demonstrates how high-level concepts are inferred from lower level specifications using semantic rules and definitions. Since we are interested in the symbolic processing of high-level interpretations and vision tasks, description logics offer a useful paradigm for modelling the different abstraction levels of our conceptual model (chapter 5).

The underlying logic and well-founded language extensions of knowledge representation formalisms give good reasoning support, which plays a crucial part of ontology in all its stages of design, maintenance and deployment. Exploiting such inference mechanisms leads to data enrichment with improved properties regarding correctness, ease of development and software reusability. However, knowledge formalisms still put us at the mercy of mathematical theories such as sets and logic. Logical computation involves regimenting arguments in ways that are often unintuitive. All sentences in logics are assertions, and reasoning based on formal logics is limited to deriving truth-values and proofs for such assertions. Hence, it is difficult to model

human reasoning that involves assumption, likelihood, belief, doubt, etc. Further, because spatial data are a concrete domain of physical and quantitative nature, we cannot divorce ourselves entirely from the mathematical definitions necessary to constrain properties, which is reflected in the classification results (chapter 7).

Overall, this approach suffers from the applicability and scalability problem of description logic languages. On the one hand, description logics live in their own realm, which is currently not interoperable with databases. The technological divide between the conceptual layer of an ontology and the data layer of a database prevent this approach from accessing data instances directly. This impedance mismatch problem requires a translation between the data and the abstract objects managed by the ontology. This task is not trivial and requires a lot of manual effort such as pre-processing and importing the data into the knowledge base. Furthermore, the use of closed world models of databases goes against the open world assumption of DLs. This meant, the thesis had to adopt methods to work around the modelling limitations by calculating spatial knowledge for the inference outside the knowledge base and asserting them through closure axioms and relevant RCC role assertions.

On the other hand, complexity barriers may still prevent DLs to become useful for larger practical applications (Neumann and Schröder, 1996). Firstly, it is unlikely that standard ABox techniques will be able to cope with large quantities, and this is especially an issue with spatial databases that consist of thousands of features given just a small geographic area. For example, the OS MasterMap coverage for the whole of the U.K. consists of over four million features. Secondly, there is a trade off between expressive power of a language and its computational complexity. OWL, for instance, is not expressive enough for some applications because its logical constructors are mainly for classes (unary predicates). There are no complex data types or built in predicates, no variables, and no higher arity predicates. Furthermore, reasoning is generally a NP hard problem, and for OWL-DL it is NExp Time-complete. This means that a solution is theoretically possible, but with problems of relevant size the solution becomes so complex that it cannot be practically achieved.

However, other techniques such as machine-learning equally suffer from the direct correlation between the complexity of the models and the complexity of the learning

techniques. Complex structures require sophisticated learning algorithms, which are mostly search procedures with exponential complexity. Furthermore, such learning programs are usually tailored to specific applications and require detailed background knowledge (Sester, 2000). Such non-standardised algorithms cannot be easily re-used. They are a black box procedure where the logic is hidden in the program, and they lack powerful inference mechanisms. Often the semantics of a given category is implicitly codified in a natural language label of a classification (Giunchiglia *et al*., 2006). In contrast, ontologies clearly specify semantics and allow reasoning over them through standard inference services. Through them, we can relate semantics to specific concepts in the data, and describe how people cognitively handle and represent these meanings.

It follows that success of this approach critically depends on the provision of efficient reasoning services to support both ontology design and deployment. This, however, is a technical limitation and not a conceptual one. To overcome these technical limitations, we need to acknowledge the current dependency between high level rules and the low level procedures that may be required to implement aspects of these rules. Machine learning, statistical analysis, and pattern recognition can provide the necessary knowledge for deriving parameters and need to be further integrated into our approach. Ontology languages are still in their infancy when it comes to modelling spatial aspects of a domain. In future, there will be increased expressive power with extended range of reasoning services, and scalability will be solved eventually (Haarslev and Möller, 2008). With active research and growing interests in the fields of description logics and ontologies, it remains to be seen how successfully the proposed conceptual framework of this thesis can be applied in future.

8.2 Potential benefits and impact

There is often a big gap between what a human user wants to do with a GIS, and the spatial concepts offered by the GIS. In particular, GIS do not sufficiently support common-sense reasoning (Egenhofer and Mark, 1995). People however perform common-sense reasoning and its outcomes make intuitive sense to them – it is reasoning that needs little explanation. To make spatial data more useful for a wider range of people, it will be necessary to incorporate people's concepts about space and time and

to mimic human thinking. A good step towards achieving this goal is the use of knowledge representation paradigms that allow the modelling of human knowledge.

The thesis demonstrates the application and functionality of a conceptual framework for modelling and deriving functional information from topographic data. The advantage of this model is its immediate appeal to common sense in terms of its conceptualisation. The definition of concepts such as semi-detached, detached or terraced house relates to common-sense knowledge and people's understanding of land use. The ability to present and to interpret spatial data in a method that is consistent with the understanding of the user leads to systems that are more flexible and will provide greater functionality in terms of cognitive spatial tasks (Hirtle, 1995). However, common-sense reasoning is difficult, and we have to account for the limitation posed by ontology languages that potentially lack the expressive power to model a specific problem or domain.

Nevertheless, the advantages that can be gained from formal knowledge representations and reasoning in general outweigh the problems, especially those resulting purely from immature technology. Working on a conceptual level not only disconnects us from the rigidness of databases, but also allows us to describe phenomena according to people's understanding. We can attach knowledge to geographical concepts and use the standard reasoning services of DL systems to derive new knowledge, which enriches data sources with new concepts. These concepts can be derived as per specification from user requirements. In addition, by having explicit semantics, ontologies enable data sharing and standardisation. For example, consider a complex spatial multi-resolution system, which has to carry out many of the above tasks. It has to derive data on demand in conformance with concepts of a specific application. By using the inference mechanisms of description logics, the system can automatically enrich its data contents and make them available to a user's specific needs. Besides the semantic generalisation of the database concepts, it also has to trigger generalisation algorithms to physically transform objects into a representation that meets its semantic conceptualisation. This can be achieved by linking concepts to algorithms, as in the work of Lüscher *et al.* (2008). For example, the semantic concept of residential area may result in aggregated topographic features denoting this concept in the data. Lastly, ontology can be used as a means to facilitate querying and providing the desired output according to theme and

scale, picking the required information from different data sources. Figure 55 illustrates the thesis in the context of such an expert generalisation system.

The computerisation of the map compilation process would save national mapping agencies incredible amounts of time and labour, and improve the consistency of their data products. Without mechanisms to formalise principles and guidelines that are well understood but difficult to exchange verbally or procedurally, manual intervention in the cartographic process not only continues to drive costs up, but impairs the quality of products. Knowledge representation formalisms such as ontology could have potential impact on operational issues for automated cartographic production and data abstraction for cartographic representation (Buttenfield and Dibble, 1995).
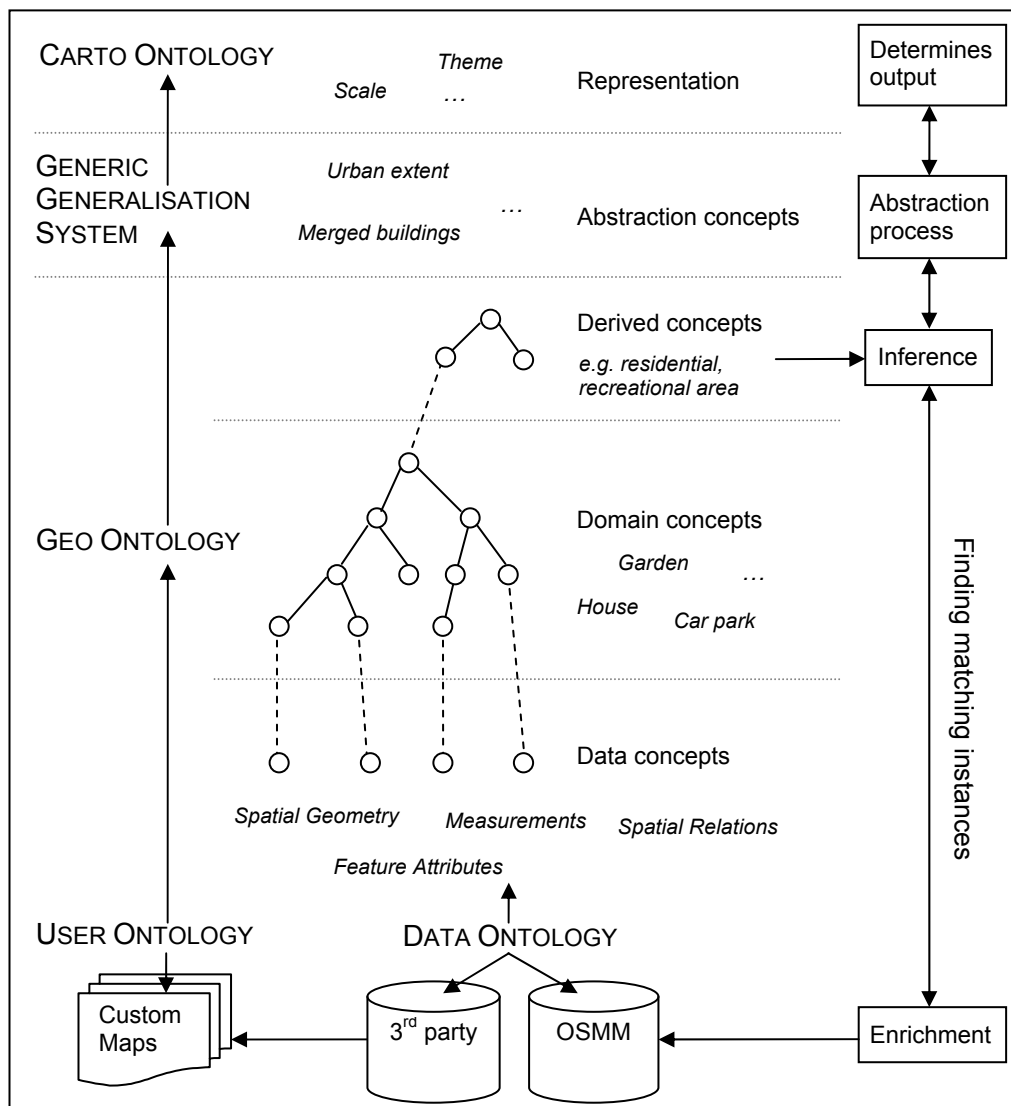
Figure 82 The thesis in the context of a generalisation system

The formalisation of human reasoning processes about spatial patterns and graphical display plays an important role in generalisation, data enrichment and other related applications. In addition, there is a pressing need within the map generalisation community to share techniques and results, which address a complex set of interoperability challenges at the technical, syntactic and semantic levels (Edwardes *et al.*, 2005). Whereas this thesis illustrates the use of semantics to reason about the functional geography at multiple levels of abstraction, the use of DL-based languages provides maximal reuse of standard components such as its reasoning services. Only the future can tell how much further research can push the boundaries of such technologies. However, I personally believe ontologies are here to stay, and that we will witness their seamless integration into our current technologies from the web to information systems. Services such as semantic query optimisation with concept languages will ease the way we search for and handle information (e.g. Buchheit *et al.*, 1994). Indeed, ontologies do not offer any magical solutions to our problems; but with ongoing research, we can foresee them to become a central part of GIScience, especially with the development of the Semantic Web and related spatial web applications.

8.3 Proposed future work

The future of GIScience relies on high-level research to build bridges to other areas, especially cognitive sciences. Spatial data models need to correspond more intuitively to people's understanding. There is a clear dichotomy between spatial data representations and higher-level geographic knowledge. People want all their data to be available from one source, to be able to share information with other people more easily, to have clear semantics defining the meaning of things represented, and to obtain generalised data on the fly. This is a long wish list, which requires more flexibility than what currently can be provided by the rigid storage models of databases.

This thesis demonstrates to some extent the potential usefulness of knowledge representation paradigms such as ontologies. Despite the current limitations of ontology languages and technology, this approach offers new avenues for exploitation. The presented ontology-based framework is flexible and should be easily adaptable to support different domains, provided the appropriate domain knowledge definitions are available. Land use categories are generated by people and are not given by the

environment per se. Hence, we have contrasting alternatives. For example, land use can be viewed as industrial, commercial, residential, and agricultural. We can reassemble pieces or categories into many alternative cognitive mappings, each one useful in a specific problem context. Therefore, the methodology needs to be transferred to other types of land uses by differentiating between classes such as industrial, recreational, and commercial, breaking them down, and integrating them in the current framework. Although, the questionnaire survey revealed that the interpretation of land use information is not straightforward in every case, other sources of knowledge, such as points of interest datasets, need to be explored for inclusion in the knowledge base to enhance the reasoning. Alternatively, the database itself can be further exploited for additional knowledge such as cartographic text labels. The proof of concept in chapter seven uses only minimal knowledge for illustration purposes.

One of the bigger challenges relates to modelling fuzzy concepts. A major weakness of this model is that it does not address the nature of uncertain knowledge, neither in the conceptual model nor in its application. The geographical domain is inherently vague with land use concepts that can overlap and physical regions whose boundaries are fuzzy. In addition, we need to formalise people's fuzzy relatedness notion in the sense of distance or proximity, and neighbourhood for distinguishing near features from far ones. Conceptualisation and formalisation of proximity and fuzziness are critical in information retrieval. Currently, the application of the proposed conceptual framework suffers from immature technology that does not incorporate probabilities efficiently for reasoning yet. This includes the immaturity of modelling the concrete spatial domains with description logics. Continuous research in these areas will eventually overcome these issues. However, it will be worthwhile to explore semantic uncertainty in the geographic domain, as for example in Ahlqvist (2004).

The high-level abstraction of the representational framework ideally should include a fourth component to accommodate fully the spatio-temporal data requirements: Time. This adds another dimension to the representation, implying a temporal type of relation on both the object and locational aspects. Time would essentially enhance a land use model that is naturally subject to change over time, and would allow incorporating the concept 'land use change'. The fundamental characteristic of relations with temporal dynamics in description logics and ontologies is therefore another major and needed

area of research for adequately modelling spatial information and making it more accessible. As Nunes (1991) notes, semantic modelling is by principle a never-ending task, just as any scientific working, but a task worthwhile to be undertaken.

Although the present beginnings seem promising in these regards, it is still much too early to utter glowing pronouncements and offer overly optimist prognosis. Ontology has yet to evolve to computationally more suitable representations. What is still required is an investment of more theoretical as well as engineering effort to data-ontology mappings (Svátek *et al.*, 2006). This task has no software support at the moment. Although the independence of symbolic logic formalisms is an advantage with respect to validity and reusability, its separate realm to databases poses a severe impediment when domain-specific properties and laws, such as dealing with space and time, must be exploited for a task. If this problem is addressed, the added value of ontologies is potentially very high. It would be unfortunate if the services of a spatial database could not be made available to the reasoning system. The spatial database should form the surrogate for a set of ABox terms, which would then allow powerful inference mechanisms over the data. The thesis demonstrates this successfully for a small sample dataset. Overall, it applies a simple, common-sense approach to enriching spatial data based on the way we interpret spatial information exploiting these kinds of formalisms. Even if our scientific communities frequently declare such formalisations as too simplistic because everyone understands them, we should instead adopt the attitude of Egenhofer and Mark (1995) that *"if it is simple and solves the problem, then it is good."*

Bibliography

ABDELMOTY, A. I., SMART, P. D., JONES, C. B., FU, G., FINCH, D. (2005) A Critical Evaluation of Ontology Languages for Geographic Information Retrieval on the Internet. *Journal of Visual Languages & Computing* **16**(4): 331-358

AGARWAL, P. (2004) Contested nature of place: knowledge mapping for resolving ontological distinctions between geographical concepts. In Egenhofer, M. J., Freksa, C. and Miller, H. J. (Eds.): *GIScience 2004*, Lecture Notes in Computer Science **3234**: 1-21, Berlin-Heidelberg: Springer-Verlag

AGARWAL, P. (2005) Ontological considerations in GIScience. *International Journal of Geographical Information Science* **19**(5): 501-536

AHLQVIST, O. (2004) A Parameterized Representation of Uncertain Conceptual Spaces. *Transactions in GIS* **8**(4): 493-514

AHLQVIST, O. (2005) Using uncertain conceptual spaces to translate between land cover categories. *International Journal of Geographical Information Science* **19**(7): 831-857

AHLQVIST, O., GAHEGAN, M. (2005) Probing the relationship between classification error and class similarity. *Photogrammetric Engineering and Remote Sensing* **71**(12): 1365-1373

ALEXANDER, C. (1965) A city is not tree. *Design* **206**: 46-55.

ANDERS, K. H., SESTER, M., FRITSCH, D. (1999) Analysis of settlement structures by graph-based clustering. *Semantische Modellierung, SMATI 99*, 41-49 Munich, Germany

ANSELIN, L. (1989) What is special about spatial data? Alternative perspectives on spatial data analysis. NCGIA, Santa Barbara, CA

ASTROVA, I. (2004) Reverse Engineering of Relational Databases to Ontologies. In Davies, J. *et al.* (Eds.): *ESWS 2004*, LNCS **3053**, pp.327-341, Springer-Verlag

ATRAN, S. (1990) *Cognitive Foundation of Natural History.* New York: Cambridge University Press.

BAADER. F., KÜSTERS, R. (2006) Nonstandard Inferences in Description Logics: The Story So Far. In Gabbay, D. M., Goncharov, S. S. and Zakharyaschev, M. (Eds.): *Mathematical Problems from Applied Logic I – Logics for the XXIst Century.* New York: Springer Science+Business Media

BAADER, F., SATTLER, U. (2001) An Overview of Tableau Algorithms for Description Logics. *Studia Logica* **69**: 5-40

BAADER, F., MCGUINNESS, D. L., NARDI, D., PATEL-SCHNEIDER, P. F. (Eds.) (2003) *The Description Logic Handbook: Theory, Implementation, and Applications.* Cambridge University Press

BAADER, F., CALVANESE, D., DE GIACOMO, G., FILLOTTRANI, P., FRANCONI, E., CUENCA GRAU, B., HORROCKS, I., KAPLUNOVA, A., LEMBO, D., LENZERINI, M., LUTZ, C., MOLLER, R., PARSIA, B., PATEL-

SCHNEIDER, P., ROSATI, R., SUNTISRIVARAPORN, B., TESSARIS, S. (2006) Formalisms for Representing Ontologies: State of the Art Survey. Technical Report TONES-D06, Tones Consortium, Available from URL: http://www.tonesproject.org/ [Last Accessed: 07/11/2008]

BADA, M., TURI, D., MCENTIRE, R., STEVENS, R. (2004) Using reasoning to guide annotation with Gene Ontology terms in GOAT. *SIGMOD Record* **33**(2): 27-32

BARKOWSKY, T., FREKSA, C. (1997) Cognitive Requirements on Making and Interpreting Maps. In Hirtle, S. and Frank, A. (Eds.): S*patial information theory: A theoretical basis for GIS, Lecture Notes in Computer Science* **1329**: 347-361 Berlin: Springer-Verlag

BARNSLEY, M. J., BARR, S. L. (1997) A graph-based structural pattern recognition system to infer land use from fine spatial resolution land cover data. *Computers, Environment and Urban Systems* **21**: 209-225

BARNSLEY, M. J., MØLLER-JENSEN, L., BARR, S. L. (2001) Inferring urban land use through spatial and structural pattern recognition. In Donnay, J. P., Barnsley, M. J., Longley, P. A. (Eds.): *Remote Sensing and Urban Analysis – GISDATA 9* London: Taylor and Francis, pp.115-144

BARR, S. L., BARNSLEY, M. J. (1997) A region-based, graph-theoretic data model for the inference of second-order thematic information from remotely-sensed images. *International Journal of Geographical Information Science* **11**(6): 555-576

BARR, S. L., BARNSLEY, M. J. (1998) Application of structural pattern-recognition techniques to infer urban land use from ordnance survey digital map data. *Proceedings of the 3rd International Conference on GeoComputation*, 17-19 September, Bristol, UK

BARR, S. L., BARNSLEY, M. J., STEEL, A. (2004) On the separability of urban land-use categories in fine spatial scale land-cover data using structural pattern recognition. *Environment and Planning B: Planning and Design* **31**: 397-418

BARWISE, J., ETCHEMENDY, J. (1999) *Language, Proof and Logic*. New York: Seven Bridges Press

BATTY, M. (2004) A New Theory of Space Syntax. Working Paper Series No. 75, London: CASA, UCL, Available from URL: http://www.casa.ucl.ac.uk/working_papers/paper75.pdf (Last Accessed: 12/10/2005)

BATTY, M., BESUSSI, E., MAAT, K., JAAP HARTS, J. (2003) Representing Multifunctional Cities: Density and Diversity in Space. Working Paper Series 71, CASA, UCL, London

BATTY, M., LONGLEY, P. (1994) *Fractal Cities*. London: Academic Press Limited

BAUER, T., STEINNOCHER, K. (2001) Per parcel land use classification in urban areas applying a rule-based technique, *GeoBIT/GIS* **6**:24-27

BCIS (2004) Guide to House Rebuilding Costs. The Building Cost Information Service, © Royal Institution of Chartered Surveyors

BECHHOFER, S., HORROCKS, I., TURI, D. (2003) Instance Store: Database support for reasoning over individuals. In Calvanese, D., De Giacomo, G. and Franconi,

E. (Eds.): *Proceedings of the 2003 Description Logic Workshop*, 5-7 September, Rome Italy

BEN-ARI, M. (1993) *Mathematic Logic for Computer Science*. Hemel Hempstead: Prentice Hall International (UK) Ltd

BENEVENTANO, D., BERGAMASCHI, S., SARTORI, C. (2003) Description logics for semantic query optimisation in object-oriented database systems. *ACM Transaction on Database Systems* **28**(1):1-50

BENJAMINS, V. R., FENSEL, D., PIERRET-GOLBREICH, C., MOTTA, E., STUDER, R., WIELINGA, B., ROUSSET, M.-C. (1999) Making Knowledge Engineering Technology Work. *Proceedings of the 9th International Conference on Software Engineering & Knowledge Engineering (SEKE'99)*, Madrid, Spain, 18-20 June

BENNETT, B. (2001) A Categorical Axiomatisation of Region-Based Geometry. *Fundamenta Informaticae* **46**: 145-158

BÉRA, R., CLARAMUNT, C. (2004) Can relative adjacency contribute to space syntax in the search for a structural logic of the city? In Egenhofer, M., Freksa, C., and Miller, H. (Eds.): *Proceedings of GIScience 2004, 3rd International Conference on Geographical Information Science,* LNCS, Springer-Verlag, Washington D.C., 20-23 October

BERENDT, B., BARKOWSKY, T. (1998) Spatial Representation with Aspect Maps. In Freksa, C., Habel, C. and Wender, K. F. (Eds.): *Spatial cognition: An interdisciplinary approach to representing an processing spatial knowledge. Lecture Notes in Artificial Intelligence* **1404**: 313-336 Berlin: Springer-Verlag

BERLIN, B., KAY, P. (1969) *Basic Color Terms: Their Universality and Growth.* Berkeley: University of California Press

BITTNER, T., STELL, J. G. (2000) Approximate qualitative spatial reasoning. *Spatial Cognition and Computation* **2**: 435-466

BLUTNER, R. (2002) Lexical semantics and pragmatics. In Hamm, F. and Zimmermann, T. E. (Eds.): *Semantics, Linguistische Berichte*, Sonderheft 10, pp.27-58

BOARD, C., TAYLOR, R. M. (1977) Perception and maps: Human factors in map design and interpretation. *Transactions of the Institute of British Geographers* **2**(1): 19-36

BOFFET, A. (2000) Creating Urban Information for Cartographic Generalisation. *Proceedings of the 9th International Symposium on Spatial Data Handling*, 10-12 August, Beijing, China pp.3b4-3b16

BOFFET, A., SERRA, R. (2000) Identification of spatial structures within urban blocks for town characterisation. *Proceedings of the 20th International Cartographic Conference (ICC)*, Beijing, China, pp.1974-1983

BOOLE, G. (1848) The Calculus of Logic. *Cambridge and Dublin Mathematical Journal* **III**: 183-198

BOYNTON, P. M., GREENHALGH, T. (2004) Selecting, designing, and developing your questionnaire. *BMJ* **328**:1312-1315

BRESCIANI, P., FRANCONI, E. (1996) Description Logics for Information Access. *Proceedings of AI*IA Workshop on Access, Extraction and Integration of Knowledge,* Napoli, September

BREY, P. (2005) The Epistemology and Ontology of Human-Computer Interaction. *Minds and Machine* **15**:383-398

BRUNER, J., GOODNOW, J., AUSTIN, A. (1956) *A Study of Thinking*. New York: Wiley.

BRUNS, H. T, EGENHOFER, M. J. (1996) Similarity of spatial scenes. In: Kraak, M.J. and Molenaar, M. (Eds.): *Advances in GIS Research II,* London: Taylor and Francis pp. 559-570

BUCHHEIT, M., JEUSFELD, M. A., NUTT, W., STAUDT, M. (1994) Subsumption between Queries to Object-Oriented Databases. *Information Systems* **19**(1): 33-54

BURROUGH, P. A., 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

BUTTENFIELD, B. P., DIBBLE, C. (1995) *NCGIA Research Initiative 8 Formalizing Cartographic Knowledge*. Scientific Report for the Specialist Meeting 24-27 October 1993, Buffalo, New York, National Center for Geographic Information and Analysis, available form URL: http://www.ncgia.ucsb.edu/Publications/Tech_Reports/95/95-15.pdf [Last accessed: 09/12/2008]

BYROM, G. (2003) Data quality and spatial cognition: the perspective of a National Mapping Agency. In *ISSDQ International Symposium on Spatial Data Quality*, Hong Kong

CALVANESE, D., DE GIACOMO, G., NARDI, D., LENZERINI, M. (2001) Reasoning in Expressive Description Logics. In Robinson, A. And Voronkov, A. (Eds.): *Handbook of Automated Reasoning*. Elsevier Science Publishers

CALVANESE, D., DE GIACOMO, G., LEMBO, D., LENZERINI, M., POGGI, A., ROSATI, R. (2006a) Linking data to ontologies: The description logic DL-Lite A. In *Proc. Of the 2nd Workshop on OWL: Experiences and Direction, OWLED 2006*

CALVANESE, D., CUENCA GRAU, B., DE GIACOMO, G., FRANCONI, E., HORROCKS, I., KAPLUNOVA, A., LEMBO, D., LENZERINI, M., LUTZ, C., MARTINENGHI, D., MÖLLER, R., ROSATI, R., TESSARIS, S., TURHAN, A.-Y. (2006b) Common Framework for Representing Ontologies. Deliverable TONES-D08

CALVANESE, D., CUENCA GRAU, B., FRANCONI, E., HORROCKS, I., KAPLUNOVA, A., LUTZ, C., MÖLLER, R., SERTKAYA, B., TESSARIS, S., TURHAN, A.-Y. (2007a) Software Tools for Ontology Design and Maintenance. Deliverable TONES-D15

CALVANESE, D., DE GIACOMO, G., LEMBO, D., LENZERINI, M., POGGI, A., ROSATI, R. (2007b) MASTRO-I: Efficient integration of relational data through DL ontologies. In Calvanese, D. *et al.* (Eds.): *Proceedings of the 2007 International Workshop on Description Logics (DL-2007),* Brixen-Bressanone, Italy, 8-10 June, CEUR-WS

CALVANESE, D., DE GIACOMO, G., GRAU, B. C., KAPLUNOVA, A., LEMBO, D., LENZERINI, M., MÖLLER, R., ROSATI, R., SATTLER, U., SERTKAYA, B., SUNTISRIVARAPORN, B., TESSARIS, S., TURHAN, A.-Y., WANDELT, S. (2007c) Ontology-Based Services: Usage Scenarios and Test Ontologies. Deliverable TONES-D14

CAMACHO-HÜBNER, E., GOLAY, F. (2007) Preliminary insights on continuity and evolution of concepts for the development of an urban morphological process ontology. *Studies in Computational Intelligence (SCI)* **61**:95-107

CARSJENS, G. J., VAN DER KNAAP, W. (2002) Strategic land-use allocation: dealing with spatial relationships and fragmentation of agriculture. *Landscape and Urban Planning* **58**: 171-179

CERBAH, F. (2008) Learning Highly Structured Semantic Repositories from Relational Databases: The RDBToOnto Tool. In Bechhofer, S. *et al.* (Eds.): *The Semantic Web: Research and Applications, ESWC 2008*, Lecture Notes in Computer Science **5021**, pp.777-781, Springer-Verlag

CHANDRASEKARAN, B., JOSEPHSON, R., BENJAMINS, V. R. (1999) What are ontologies, and why do we need them? *IEEE Intelligent Systems* **14**(1): 20-26

CHAUDHRY, O. (2008) Modelling geographic phenomena at multiple levels of detail: A model generalisation approach based on aggregation. Ph.D. dissertation, University of Edinburgh

CHAUDHRY, O., MACKANESS, W. A. (2007) Utilising partonomic information in the creation of hierarchical geographies. *10th ICA Workshop on Generalisation and Multiple Representation*, 2-3 August 2007, Moscow

CHIERCHIA, G. (1999) Linguistics and Language. In Wilson, R.A. and Keil, F.C. (Eds.): *The MIT Encyclopedia of the Cognitive Sciences*, pp. xv-xxxvii, Massachusetts: The MIT Press

CHOU, Y.-H. (1995) Spatial pattern and spatial autocorrelation. In Frank, A. U. and Kuhn, W. (Eds.): *Spatial Information Theory: A theoretical basis for GIS. Proceeding of COSIT'95*, Semmering, Austria, September 21-23, LNCS, Vol. 988, pp.365-376, Berlin: Springer-Verlag

CHRISTOPHE, S., RUAS, A. (2002) Detecting Building Alignments for Generalisation Purposes. *Symposium on Geospatial Theory, Processing and Applications,* Ottawa

CLAWSON, M., STEWART, C. L. (1965) *Land use information – A critical survey of US statistics including possibilities for greater uniformity.* Baltimore: Johns Hopkins Press

CLEMENTINI, E., DI FELICE, P., HERNÁNDEZ, D. (1997) Qualitative representation of positional information. *Artificial Intelligence* **95**: 317-356

COHN, A. G., HAZARIKA, S. M. (2001) Qualitative Spatial Representation and Reasoning: An overview, *Fundamenta Informaticae* **46**(1-2): 1-29

COHN, A. G., BENNETT, B., GOODAY, J., MARK GOTTS, N. (1997) Qualitative Spatial Representation and Reasoning with the Region Connection Calculus. *GeoInformatica* **1**: 275-316

COMBER, A., FISHER, P., WADSWORTH, R. (2005a) What is land cover? *Environment and Planning B: Planning and Design* **32**:199-209

COMBER, A., FISHER, P., WADSWORTH, R. (2005b) You know what land cover is but does anyone else? ...an investigation into semantic and ontological confusion. *International Journal of Remote Sensing* **26**(1): 223-228

COMBER, A., FISHER, P., WADSWORTH, R. (2005c) Comparing and Combining Different Expert Relations of How Land Cover Ontologies Relate. In Fisher, P. (Ed.): *Developments in Spatial Data Handling,* 11th International Symposium on Spatial Data Handling, pp.573-583, Springer-Verlag

CONROY DALTON, R., KIRSAN, C. (2005) Graph isomorphism and genotypical houses. In: The Fifth International Space Syntax Symposium, 13-17 June 2005, Delft, Holland.

CUENCA GRAU, B., HORROCKS, I., MOTIK, B., PARSIA, B., PATEL-SCHNEIDER, P., SATTLER, U. (2008) OWL2: The Next Step for OWL. *Journal of Web Semantics* **6**(4): 309-322

CUENCA GRAU, B., HORROCKS, I., PARSIA, B., PATEL-SCHNEIDER, P., SATTLER, U. (2006) Next Steps for OWL. *Proc. of the Second OWL Experiences and Directions Workshop*, volume **216** of CEUR

CUTINI, V. (2003) Lines and squares: Towards a configurational approach to the morphology of open spaces, *Proceedings of the 4th International Space Syntax Symposium*, London

CUTINI, V., PETRI, M., SANTUCCI, A. (2004) From Axial Maps to Mark Point Parameter Analysis (Ma.P.P.A.) – A GIS Implemented Method to Automate Configurational Analysis, *ICCSA* (2) pp.1107-1116

D'AGOSTINO, M. (1992) Are Tableaux an Improvement on Truth-Tables? *Journal of Logic, Language, and Information* **1**: 235-252

DAVIES, C. (2006) Getting from cognition to collection: data provision for usable models. *Cognitive Approach to Modelling Environments (CAME) Workshop, GIScience 2006 Conference*, Münster, Germany, 20 September

DAVIES, C., MEDYCKYJ-SCOTT, D. (1996) GIS users observed. *International Journal of Geographical Information Systems* **10**(4): 363-384

DAVIES, C., WOOD, L., FOUNTAIN, L. (2005) User-centred GI: hearing the voice of the customer. In *Annual conference of the Association for Geographic Information AGI'05, People, Places and Partnerships,* November 8-10, London

DAVIES, J., FENSEL, D., VAN HARMELEN, F. (Eds.) (2003) *Towards the semantic web: Ontology-driven knowledge management.* Chichester: John Wiley & Sons Ltd

DAVIS, R., SHROBE, H., SZOLOVITS, P. (1993) What is a Knowledge Representation? *AI Magazine* (Spring 1993): 17-33

DE ALMEIDA J.-P. (2007). "A Graph-based Technique for Analysis and Visualisation of Higher Order Urban Topology". PhD thesis (unpublished). Department of Geomatic Engineering, University College London (London, UK).

DEY, A. K., ABOWD, G. D. (1999) Towards a Better Understanding of Context and Context-Awareness. In Gellersen, H.-W. (Ed): *Proceedings of the 1st*

International Symposium on Handheld and Ubiquitous Computing. Karlsruhe, Germany. Lecture Notes in Computer Science **1707**: 304-337, Berlin: Springer-Verlag

DOLBEAR, G., GOODWIN, J. (2007) Expressing Relational Data as RDF. *W3C Workshop on RDF Access to Relational Databases*, Cambridge, MA, USA, 25-26 October

DOLBEAR, C., HART, G., GOODWIN, J. (2007) From Theory to Query: Using ontologies to make explicit imprecise spatial relationships for database querying. *COSIT 2007*, Melbourne, Australia, 19-23 September

DONINI, F. M., LENZERINI, M., NARDI, D., SCHAERF, A. (1996) Reasoning in Description Logics. In Brewka, G. (Ed.): *Principles of Knowledge Representation and Reasoning, Studies in Logic, Language and Information.* CLSI Publications, pp.193-238

DOWNS, R.M., STEA, D. (1977) *Maps in Minds - Reflections on Cognitive Mapping.* New York: Harper & Row

DUCHÊNE, C., BARD, S., BARILLOT, X., RUAS, A., TRÉVISAN, J., HOLZAPFEL, F. (2003) Quantitative and qualitative description of building orientation. *Fifth Workshop on Progress in Automated Map Generalisation,* ICA commission on Map Generalisation, Paris, France, April

DUINEVELD, A., STOTER, R., WEIDEN, M. R., KENEPA, B., BENJAMINS, V. R. (1999) Wonder Tools? A Comparative Study of Ontological Engineering Tools. *12th International Workshop on Knowledge Acquisition, Modeling and Management (KAW99)*, 16-21 October, Banff, Canada

DUTTON, G., EDWARDES, A. (2006) Ontological Modeling of Geographical Relationships for Map Generalization. *Workshop of the ICA Commission on Map Generalisation and Multiple Representation*, Portland, USA, 25th June

EDWARDES, A., BURGHARDT, D., NEUN, M. (2005) Interoperability in Map Generalisation Research. In *Proceedings of the International Symposium on Generalisation of Information, ISGI 2005*, 14-16 September, Berlin

EGENHOFER, M. (1989) A Formal Definition of Binary Topological Relationships, *Lecture Notes in Computer Science* **367** pp.457-472

EGENHOFER, M. (1993) What's special about spatial? Database requirements for vehicle navigation in geographic space. *Sigmond Record* **22**:398-402

EGENHOFER, M. J., AL-TAHA, K. K. (1992) Reasoning about Gradual Changes of Topological Relationships. In Frank, A., Campari, I. and Formentini, U. (Eds.): *Lecture Notes in Computer Science,* Berlin, Germany: Springer-Verlag, vol. 639, Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, pp.169-219

EGENHOFER, M., HERRING, J. (1991) Categorizing Binary Topological Relationships between Regions, Lines, and Points in Geographic Databases. In Egenhofer, M., Herring, J., Smith, T., and Park, K. (Eds.): *A Framework for the Definition of Topological Relationships and an Algebraic Approach to Spatial Reasoning within this Framework, NCGIA Technical Report 91-7,* Santa Barbara, CA: National Center for Geographic Information and Analysis

EGENHOFER, M., HERRING, J. R., SMITH, T., PARK, K. K. (1991) A Framework for the Definition of Topological Relationships and an Algebraic Approach to Spatial Reasoning within this Framework, NCGIA Technical Report 91-7, Santa Barbara, CA: National Center for Geographic Information and Analysis

EGENHOFER, M. J., MARK, D. M. (1995) Naive Geography. COSIT'95, Semmering, Austria, in Frank, A. and Kuhn, W. (Eds.): *Lecture Notes in Computer Science* **988**: 1-15 Springer-Verlag

EL-GHALAYINI, H., ODEH, M., MCCLATCHEY, R., SOLOMONIDES, T. (2005) Reverse engineering ontology to conceptual data models. In Hamza M. H. (Ed.): *IASTED International Conference on Databases and Applications*, 23rd Multi-Conference on Applied Informatics, Innsbruck, Austria, 14-16 February, pp.222-227

ENGEL, D., BERTEL, S., BARKOWSKY, T. (2005) Spatial principles in control of focus in reasoning with mental representation, images, and diagrams. In Freksa, C. *et al.* (Eds.): *Spatial Cognition IV*, LNAI **3343**, pp.181-203, Springer-Verlag

ESPOSITO, F., LANZA, A., MALERBA, D., SEMERARO, G. (1997) Machine learning for map interpretation: An intelligent tool for environmental planning. *Applied Artificial Intelligence* **11**: 673-696

EVERMANN, J., WAND, Y. (2005) Ontology based object-oriented domain modelling: fundamental concepts. *Requirements Engineering* **10**: 146-160

FABRIKANT, S. I., BUTTENFIELD, B. P. (1997) Envisioning user access to a large data archive. *Proceedings GIS/LIS'97*, Cincinnati, Ohio, 28-30 October, pp.686-691

FAGIN, R., HALPERN, J. Y., MOSES, Y., VARDI, M. Y. (2003) *Reasoning About Knowledge*. Massachusetts: The MIT Press

FAYYAD, U., PIATETSKY-SHAPIRO, G., SMYTH, P. (1996) Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proc. 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)*, AAAI Press, Menlo Park, CA, pp.82-88

FELLBAUM, C. (Ed.) (1998*) WordNet – an electronic lexical database.* Massachusetts: The MIT Press

FENG, C.-C., FLEWELLING, D. M. (2004) Assessment of semantic similarity between land use/land cover classification systems. *Computers, Environment and Urban Systems* **28**:229-246

FONSECA, F. (2001) *Ontology-driven Geographic Information Systems*. PhD Thesis, University of Maine

FONSECA, F., EGENHOFER, M. (1999) Ontology-Driven Geographic Information Systems. In Bauzer Medeiros (Ed.): *7th ACM Symposium on Advances in Geographic Information Systems*, Kansas City, MO, November

FONSECA, F., DAVIS, C., CÂMARA, G. (2003) Bridging Ontologies and Conceptual Schemas in Geographic Information Integration. *GeoInformatica* **7**(4): 355-378

FONSECA, F., EGENHOFER, M., AGOURIS, P. (2002) Using Ontologies for Integrated Geographic Information Systems. *Transactions in GIS* **6**(3): 231-257

FONSECA, F., EGENHOFER, M., DAVIS JR., C. A., BORGES, K. A. V. (2000) Ontologies and Knowledge Sharing in Urban GIS. *Computer, Environment and Urban Systems* **24**(3): 232-251

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

FRANK, A. U. (2001) Tiers of ontology and consistency constraints in geographical information systems. *International Journal of Geographical Information Science* **15**(7): 667-6678

FRANK, A. U. (2003) Ontology for Spatio-temporal Databases. In T. Sellis *et al.* (Eds.): *Spatio-Temporal Databases,* chapter 2, Lecture Notes in Computer Science 2520, pp.9-77 Berlin: Springer-Verlag

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

FREKSA, C. (1991) Conceptual Neighborhood and its role in temporal and spatial reasoning. In Singh, M. and Travé-Massuyès, L. (Eds.): *Proceedings of the IMACS Workshop on Decision Support Systems and Qualitative Reasoning*, Amsterdam, Netherlands, pp.181-187

FREKSA, C. (1994) Fuzzy Systems in AI. In Kruse, R., Gebhardt, J., Palm, R. (Eds.): *Fuzzy systems in computer science*. Braunschweig: Vieweg, pp.155-169

FREKSA, C., BARKOWSKY, T. (1995) On the relation between spatial concepts and geographic objects. In Burrough, P. and Frank, A. U. (Eds.): *Geographic Objects with Undetermined Boundaries*. London: Taylor & Francis, pp.99-121

FREUNDSCHUH, S. M., EGENHOFER, M. J. (1997) Human Conceptions of Spaces: Implications for Geographic Information Systems. *Transactions in GIS* **2**(4): 361-375

GAFFURI, J., TRÉVISAN, J. (2004) Role of urban patterns for building generalisation: An application of AGENT. *ICA Workshop on Generalisation and Multiple Representation*, 20-21 August, Leicester, England

GALE, N., GOLLEDGE, R. G. (1982) On the Subjective Partitioning of Space. *Annals of the Association of American Geographers* **72**(1): 60-67

GALTON, A. (1999) The Mereotopology of Discrete Space. In Freksa, C. and Mark, D. M. (Eds.): *COSIT'99*, LNCS **1661**, pp.251-266, Springer-Verlag

GAŠEVIĆ, D., DJURIĆ, D., DEVEDŽIĆ, V. (2006) *Model Driven Architecture and Ontology Development.* Berlin: Springer

GÄRDENFORS, P. (2000) *Conceptual Spaces: On the Geometry of Thought.* Cambridge, MA: MIT Press

GÄRLING, T., GOLLEDGE, R. (1987) Environmental perception and cognition. In Zube, E. and Moore, G. (Eds.): *Advances in Environment, Behaviour, and Design,* Vol. 2, New York: Plenum, pp. 203-236

GEHRKE, J., HOMMEL, B. (1998) The impact of exogenous factors on spatial coding in perception and memory. In Freksa, C., Habel, C. and Wender, K. F. (Eds.): *Spatial Cognition An Interdisciplinary Approach to Representing and*

Processing Spatial Knowledge, Lecture Notes in Artificial Intelligence **1404**: 63-77 Berlin: Springer-Verlag

GENNARI, J., MUSEN, M. A., FERGERSON, R. W., GROSSO, W. E., CRUBEZY, M., ERIKSSON, H., NOY, N. F., TU., S. W. (2003) The evolution of Protégé: An environment for knowledge-based systems development. *International Journal of Human-Computer Studies* **58**(1): 89-123

GEOINFORMATION GROUP (2004) *CR Land Use Product Sheet*. Online product sheet, available from URL: http://www.citiesrevealed.com/aware/aware_fr.htm (last accessed 28/07/2006), Cambridge, GeoInformation Group

GIBSON, J. J. (1979) *The Ecological Approach to Visual Perception*. Houghton Mifflin, Boston

GIUNCHIGLIA, F., MARCHESE, M., ZAIHRAYEU, I. (2006) Encoding Classification into Lightweight Ontologies. In Sure, Y. and Domingue, J. (Eds.): *ESWC2006*, LNCS **4011**, pp.80-94, Springer-Verlag

GOODCHILD, M. F. (2004) GIScience: geography, form, and process. *Annals of the Association of American Geographers* **94**(4): 709-714

GOODWIN, J. (2005) Experiences of using OWL at the Ordnance Survey. *OWL: Experiences and Directions (OWLED Workshop)*, 11-12 November, Galway, Ireland

GOLLEDGE, R. (1982) Fundamental conflicts and the search for geographical knowledge. In Gould, P. and Olsson, G. (Eds.): *A search for common ground.* London: Pion, pp.11-23

GOLLEDGE, R. G. (1992) *Do People Understand Spatial Components: The Case of First-Order Primitives.* Working Paper UCTC No.211, The University of California at Berkeley, Transportation Centre. Presented at the International GIS Conference, Pisa, Italy

GOLLEDGE, R. G. (2002) The Nature of Geographic Knowledge. *Annals of the Association of American Geographers* **92**(1): 1-14

GREEN, J., THOROGOOD, N. (2004) Q*ualitative Methods for Health Research.* London: Sage Publications Ltd.

GRITTENS, J. S. (1969) Forming Impressions of an Unfamiliar City: A Comparative Study of Aesthetic and Scientific Knowing. MA Thesis, Clark University, Worcester, Massachusetts.

GRUBER, T. R. (1993) Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In Guarino, N. and Poli, R. (Eds.): *Formal Ontology in Conceptual Analysis and Knowledge Representation.* Kluwer Academic Publishers

GRÜNINGER, F., FOX, S. M. (1995) Methodology for the Design and Evaluation of Ontologies. *Proceedings of the Workshop on Basic Ontological issues in Knowledge Sharing, IJCAI-95*, April 13, Montreal

GRÜTTER, R., BAUER-MESSMER, B. (2007a) Towards Spatial Reasoning in the Semantic Web: A Hybrid Knowledge Representation System Architecture. In Fabrikant, S. and Wachowicz, M. (Eds.): *The European Information Society:*

Leading the Way with Geo-information. Lecture Notes in Geoinformation and Cartography, pp. 349-364

GRÜTTER, R., BAUER-MESSMER, B. (2007b) Combining OWL with RCC for Spatioterminological Reasoning on Environmental Data. *In Third International Workshop OWL: Experiences and Directions, OWLED'07*, June

GRÜTTER, R., SCHARRENBACH, T., BAUER-MESSMER, B. (2008) Improving an RCC-Derived Geospatial Approximation by OWL Axioms. In Sheth, A. *et al.* (Eds.): *The Semantic Web – ISWC 2008*, 7th International Semantic Web Conference, Karlsruhe, Germany, 26-30 October, Lecture Notes in Computer Science **5318**, pp. 293-306, Springer-Verlag

GUARINO, N. (1995) Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies* **43**: 625-640

GUARINO, N. (1998) Formal Ontology and Information Systems. In Guarino, N. (Ed.): *Formal Ontology in Information systems, Proceedings of FOIS'98*, Trento, Ital, 6-8 June, IOS Press, pp.3-15

GUARINO, N., GIARETTA, P. (1995) Ontologies and knowledge bases: Towards a terminological clarification. In Mars, N. (Ed.): *Towards Very Large Knowledge Bases: Knowledge Building and Knowledge Sharing.* Amsterdam: IOS Press, pp. 25-32

GUO, D., PEUQUET, D., GAHEGAN, M. (2002) Opening the Black Box: Interactive Hierarchical Clustering for Multivariate Spatial Patterns. *The Tenth ACM International Symposium on Advances in Geographic Information Systems,* McLean, VA, November 8-9, pp.131-136

GUSTAFSON, E. J. (1998) Quantifying landscape spatial pattern: What is state of the art? *Ecosystems* **1**(2): 143-156

GÜTING, R. H. (1994) An Introduction to Spatial Database Systems. *VLDB Journal* **3**: 357-399

HAARSLEV, V., MÖLLER, R. (1997) SBox: A Qualitative Spatial Reasoner – Progress Report. In Ironi, L. (Ed.): *Proceedings of the 11th International Workshop on Qualitative Reasoning*, Cortona, Tuscany, Italy, 3-6 June, pp.105-113

HAARSLEV, V., MÖLLER, R. (2000) Expressive ABox Reasoning with Number Restrictions, Role Hierarchies, and Transitively Closed Roles. In Cohn, A. G., Giunchiglia, F. and Selman, B. (Eds.): *International Conference on Principles of Knowledge Representation and Reasoning (KR'2000)*, April

HAARSLEV, V., MÖLLER, R. (2001) High performance reasoning with very large knowledge bases: A practical case study. In Nebel, B., Levesque, H. (Eds.): *International Joint Conference on Artificial Intelligence (IJCAI'2001)*, 4-10 August, Seattle, Washington, USA, Morgan-Kaufmann

HAARSLEV, V., MÖLLER, R. (2008) On the Scalability of Description Logic Instance Retrieval. *Journal of Automated Reasoning* **41**(2): 99-142

HAARSLEV, V., MÖLLER, R., SCHRÖDER, C. (1994) Combining Spatial and Terminological Reasoning. In Nebel, B. and Dreschler-Fischer, L. (Eds.): *KI-94, Advances in Artificial Intelligence – Proc. 18th German Annual Conference on*

Artificial Intelligence, Saarbrücken, 18-23 September, Lecture Notes in Artificial Intelligence **861**, pp.142-153, Springer-Verlag

HAGGETT, P., CHORLEY, R. J. (1969) *Network Analysis in Geography*. London: Edward Arnold

HAKLAY, M. (2006) Usability dimensions in collaborative GIS. In Dragicevic, S. and Balram, S. (Eds.): *Collaborative Geographical Information Systems*. London: Idea Group, pp.24-42

HAKLAY, M., ZAFIRI, A. (2008) Usability Engineering for GIS: Learning from a Screenshot. *The Cartographic Journal* **45**(2): 87-97

HALL, M. (2006) Automatisierte semantische Harmonisierung von Landnutzungsdaten. In Strobl, J., Blaschke, T. und Griesebner, G. (Hrsg.): *Angewandte Geoinformatik 2006*. Beiträge zum 18. AGIT-Symposium, Salzburg, Herbert Wichman Verlag, Heidelberg

HANSEN, H. (2003) A fuzzy logic approach to urban land-use mapping. In Virrantaus, K. and Tweite, H. (Eds.): *ScanGIS2003 – The 9th Scandinavian Research Conference on Geographical Information Science*, 4-6 June, Espoo, Finland, pp.41-56

HARALICK, R. M., SHAPIRO, L. G. (1992) *Computer and Robot Vision.* Addison-Wesley, Vol. I & II

HARDING, J. (2003) *Understanding how users view the world – from spatial cognition to geographic information.* MS PowerPoint presentation, Ordnance Survey, Southampton

HARDING, J., DAVIES, C. (2004) Concepts, Semantics and Geographies. *Workshop on the Potential of Cognitive Semantics for Ontologies, FOIS 2004*

HARRIS, R. J., LONGLEY, P. A. (2000) New data and approaches for urban analysis: Modelling residential densities. *Transactions in GIS* **4**(3): 217-234

HARRISON, A. R. (2000) The National Land Use Database: developing a framework for spatial referencing and classification of land use features. *Proceedings of the AGI Conference at GIS 2000*, Olympia, London, pp.5.6.1-5.6.8

HARRISON, A. R. (2002) Extending the dimensionality of OS MasterMap^TM: land use and land cover, *Proceedings of the AGI conference at GIS 2002*, Olympia, London, pp.4.1.1 - 4.1.12

HARRISON, A. R., GARLAND, B. (2001) The National Land Use Database: building a new national baseline data of urban and rural land use, *Proceedings of the AGI Conference at GIS 2001*, Olympia, London, pp.2.5.1 - 2.5.11

HART, G. (2007) Ontology Database Connectivity – Field Definition. Ordnance Survey Research, Internal report, Southampton, ©Crown copyright

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

HARTOG, J., HOLTROP, B. T., GUNST, M. E., OOSTERBROEK, E.-P. (1999) Interpretation of Geographic Vector-Data in Practice. In Chhabra, A.K. and Dori, D. (Eds.): *GREC'99, Lecture Notes in Computer Science* **1941**: 50-57 Berlin Springer-Verlag

HARTZ, J., NEUMANN, B. (2007) Learning a Knowledge Base of Ontological Concepts for High-Level Scene Interpretation. *Proceedings of the Sixth International Conference on Machine Learning and Applications, ICMLA'07,* Cincinnati, Ohio, 13-15 December

HEDMAN, S. (2006) *A First Course in Logic - An Introduction to Model Theory, Proof Theory, Computability, and Complexity.* Oxford: Oxford University Press

HEINZLE, F., SESTER, M. (2004) Derivation of implicit information from spatial data sets with data mining. *International Archives of Photogrammetry and Remote Sensing* **35**:335-340

HEINZLE, F., SESTER, M., ANDERS, K.-H. (2005) Graph based approach for recognition of patterns and implicit information in road networks. *Proceedings of ICA Conference*, 9-16 August, A Coruña (Spain)

HELD, R., REKOSH, J. (1963) Motor-sensory feedback and the geometry of visual space. *Science* **141**: 722-723

HERETH, J., STUMME, G., WILLE, R., WILLE, U. (2000) Conceptual knowledge discovery and data analysis. In Mineau, G. and Ganter, B. (Eds.): *International Conference on Conceptual Structures, LNCS 1867*, Springer-Verlag, pp.421-437

HILLIER, B. (1996) *Space is the machine: a configurational theory of architecture.* New York: Cambridge University Press

HILLIER, B., HANSON, J. (1984) *The Social Logic of Space.* Cambridge University Press, Cambridge

HILLIER, B., LEAMAN, A., STANSALL, P., BEDFORD, M. (1976) Space Syntax, *Environment and Planning B* **3**: 147-185

HILLIER, B., PENN, A. (2004) Rejoinder to Carlo Ratti. *Environment and Planning B: Planning and Design* **31**: 501-511

HIRTLE, S. C. (1995) Representational Structures for Cognitive Space: Trees, Ordered Trees and Semi-Lattices. In Frank, A. U. and Kuhn, W. (Eds.): *Spatial Information Theory: A Theoretical Basis for GIS*, International Conference COSIT'95, Semmering, Austria, 21-23 September, Lecture Notes in Computer Science **988**: 327-340, Springer-Verlag

HOCKENBERRY, M. C. (2004) Grounding for a Computational Model of Place. Thesis in Master of Science. Massachusetts Institute of Technology, Media Arts and Sciences

HOFER, B. K., PINTRICH, P. R. (1997) The development of epistemological theories: beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research* **67**(1): 88-140

HOLYOAK, K. J. (1999) Psychology. In Wilson, R.A. and Keil, F.C. (Eds.): *The MIT Encyclopedia of the Cognitive Sciences*, pp. xxxvii-xlix, Massachusetts: The MIT Press

HORNSBY, K., EGENHOFER, M. J. (2000) Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographical Information Science* **14**(3): 207-224

HORRIDGE, M., DRUMMOND, N., GOODWIN, J., RECTOR, A., STEVENS, R., WANG, H. H. (2006) The Manchester OWL Syntax. *Proceedings of the OWL Experiences and Direction Workshop (OWLED'06)*, Athens, Georgia

HORROCKS, I. (2005a) OWL: A Description Logic Based Ontology Language. In van Beck, P. (Ed.): *CP 2005*, LNCS **3709**, pp5-8, Springer-Verlag

HORROCKS, I. (2005b) Description Logics in Ontology Applications. In Beckert, B. (Ed.): *Automated reasoning with analytic tableaux and related methods*. International Conference, TABLEAUX 2005, Koblenz, Germany, 14-17 September, Lecture Notes in Computer Science **3702**, pp.2-13 Springer-Verlag

HORROCKS, I. (2006) OWL: A Description Logic Based Ontology Language. *Seminar at the Centre for Intelligent Systems and their Applications (CISA)*, Edinburgh, Scotland, 30th March 2006

HORROCKS, I., PATEL-SCHNEIDER, P. F. (2003) Three Theses of Representation in the Semantic Web. *Proceedings of the 12th International World Wide Web Conference, WWW 2003*, 20-24 May, Budapest, Hungary, pp. 39-47

HORROCKS, I., PATEL-SCHNEIDER, P. F., VAN HARMELEN, F. (2003) From SHIQ and RDF to OWL: The Making of a Web Ontology Language. *Journal of Web Semantics* **1**(1): 7–26

HORROCKS, I., SATTLER, U. (2005) A tableaux decision procedure for SHOIQ. *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI 05)*

HOTZ, L., NEUMANN, B. (2005) Scene Interpretation as a Configuration Task. *Künstliche Intelligenz* **3**:59-65

HUANG, Y., DIMITROVA, V., AGARWAL, P. (2005) Detecting mismatches between a user's and a system's conceptualizations to empower user modelling for the Semantic Web. *PersWeb05 workshop, User Modelling Conference*, 28-29 July, Edinburgh

HULL, R., KING, R. (1987) Semantic Database Modeling: Survey, Applications, and Research Issues. *ACM Computing Surveys* **19**(3): 201-260

HUSSAIN, M., DAVES, C., BARR, R. (2007) Classifying Buildings Automatically: A Methodology. In Winstanley, A. (Ed.): *GISRUK 2007 Proceedings of the Geographical Information Science Research UK Conference*, 11-13th April, National Centre for Geocomputation, National University of Ireland Maynooth

HUTCHINS, E. (1995) *Cognition in the Wild.* Cambridge, M.A.: MIT Press.

ISLI, A., HAARSLEV, V., MÖLLER, R. (2001) Combining cardinal direction relations and relative orientation relations in qualitative spatial reasoning. Technical report fbi-hh-m-304/01, LKI, University of Hamburg, Computer Science Department

JAHN, G., JOHNONS-LEIRD, P. N., KNAUFF, M. (2005) Reasoning about consistency with spatial mental models: hidden and obvious indeterminacy in spatial descriptions. In Freksa, C. *et al.* (Eds.): *Spatial Cognition IV, Lecture Notes in Artificial Intelligence* **3343**: 165-180 Berlin: Springer-Verlag

JAKULIN, A., MLADENIĆ, D. (2005) Ontology grounding. In Markič *et al.*, (Eds.): *Proceedings of the 8th International multi-conference Information Society IS-2005*, 17 October, Ljubljana, Slovenia, pp.174-177

JEROME, J. (2003) Thinking beyond the data: Creating knowledge from geospatial chaos. In *Proceedings of the AGI conference at GeoSolutions 2003*, pp. 1-3

JIANG, B., CLARAMUNT, C. (2004) A Structural Approach to the Model Generalisation of an Urban Street Network. *GeoInformatica* **8**(2): 157-171

JONES, D. M., BENCH-CAPON, T., VISSER, P. (1998) Methodologies for Ontology Development. In Cuena, J. (Ed.): *Proc. IT&KNOWS Conference, XV IFIP World Computer Congress*, Budapest, August, pp.62-75, Chapman-Hall

JONES, C. B., WARE, J. M. (2005) Map Generalisation in the Web Age. *International Journal of Geographical Information Science* **19**(8-9): 859-870

JORDAN, T., RAUBAL, M., GARTRELL, B., EGENHOFER, M. J. (1998) An affordance-based model for place in GIS. In Poiker, T. and Chrisman, N. (Eds.): *Eighth International Symposium on Spatial Data Handling*, Vancouver, Canada, pp.98-109

KASEMSOOK, A. (2003) Spatial and functional differentiation: A symbiotic and systematic relationship, *Proceedings of the 4th International Space Syntax Symposium,* London

KASHYAP, V. (1999) Design and creation of ontologies for environmental information retrieval. *Proceedings of the 12th Workshop on Knowledge Acquisition, Modelling and Management*, 2-6 October, Banff, Alberta, Canada

KASHYAP, V., SHETH, A. (1996) Semantic and Schematic Similarities between Database Objects: A Context-based approach. *VLDB Journal* **5**(4):276-304

KATZ, Y., CUENCA GRAU, B. (2005) Representing Qualitative Spatial Information in OWL-DL. *Proceedings of OWL: Experiences and Directions*, Galway, Ireland

KATZ, J. J., FODOR, J. A. (1963) The structure of a semantic theory. *Language* **39**(2): 170-210

KAVOURAS, M. (2003) A unified ontological framework for semantic integration. *Proceedings of the International Workshop on Next Generation Geospatial Information*, 19-21 October, Cambridge (Boston), Massachusetts, USA

KEITA, A., LAURINI, R., ROUSSEY, C., ZIMMERMAN, M. (2004) Towards an Ontology for Urban Planning: The Towntology Project. *Proceedings of the 24th UDMS Symposium,* 27-29 October, Chioggia

KEßLER, C. (2006) Conceptual Spaces for Data Descriptions. *Proceedings of Workshop on the Cognitive Approach to Modeling Environments (CAME)*, GIScience 2006, Münster, Germany, SFB/TR 8 Report No. 009-08/2006: 29-35

KITZINGER, J. (1995) Introducing focus groups. *BMJ* **311**:299-302

KLEIN, R., BUCHHEIT, M., NUTT, W. (1994) Configuration as model construction: The constructive problem solving approach. In Sudweeks, F. and Gero, J. (Eds.): *Proceedings of the 4th International Conference on Artificial Intelligence in Design*, Kluwer, Dordrecht

KLIEN, E., EINSPANIER, U., LUTZ, M., HÜBNER, S. (2004) An architecture for ontology-based discovery and retrieval of geographic information. In: *Proceedings 7th Conference on Geographic Information Science, AGILE 2004,* Heraklion, Greece.

KLIEN, E., 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, *Lecture Notes in Computer Science* **3693**: 133-148, Berlin Heidelberg, Springer-Verlag

KLIEN, E., PROBST, F. (2005) Requirements for Geospatial Ontology Engineering. In *Proceedings of the 8th AGILE Conference on Geographic Information Science*, Estoril, Portugal

KLINOV, P. (2008) Pronto: A Non-monotonic Probabilistic Description Logic Reasoner. In Bechhofer, S., Hauswirth, M., Hoffman, J., Koubarakis, M. (Eds.): *The Semantic Web: Research and Applications, 5th European Semantic Web Conference, ESWC 2008*, Tenerife, Canary Islands, Spain, 1-5 June, Lecture Notes in Computer Science **5021**, pp. 822-826

KOCH, H., PAKZAD, K., TÖNJES, R. (1997) Knowledge Based Interpretation of Aerial Images and Maps Using a Digital Landscape Model as Partial Interpretation. *Workshop on Semantic Modeling for the Acquisition of Topographic Information from Images and Maps, SMATI'97*, 21-23 May, Bonn, Germany

KOKLA, M., KAVOURAS, M. (2001) Fusion of top-level and geographical domain ontologies based on context formation and complementarity. *International Journal of Geographical Information Science* **15**(7): 679-687

KOLA, J., RECTOR, A. (2007) Importing Spreadsheet data into Protégé: Spreadsheet plug-in. *10th International Protégé Conference*, Budapest, Hungary, 15-18 July

KOLÁČNÝ, A. (1969) Cartographic Information − a Fundamental Concept and Term in Modern Cartography. *Cartographic Journal* **6**: 47-9

KOSSLYN, S. M. (1978) Imagery and Internal Representation. In Rosch, E. and Lloyd, B. B. (Eds.): *Cognition and Categorization*. New Jersey: Lawrence Erlbaum Associates

KOPERSKI, K., HAN, J. (1995) Discovery of Spatial Association Rules in Geographic Information Databases, in Egenhofer, M.J. and Herring, J.R. (Eds.): *Advances in Spatial Databeses'95, Lecture Notes in Computer Science*, Vol. 951, Heidelberg: Springer-Verlag, pp.47-66

KOPERSKI, K., HAN, J., STEFANOVIC, N. (1998) An Efficient Two-Step Method for Classification of Spatial Data. *Proc. 8th International Symposium on Spatial Data Handling, SDH'98*, 11-15 July, Vancouver, Canada, pp.45-54

KOVACS, K., DOLBEAR, C., HART, G., GOODWIN, J, MIZEN, H. (2006) A methodology for building conceptual domain ontologies. Ordnance Survey Research, Southampton

KUHN, W. (2004a) Why Information Science needs Cognitive Semantics – and what it has to offer in return. *Workshop on the Potential of Cognitive Semantics for Ontologies, FOIS 2004*, 4-6 November, Torino, Italy

KUHN, W. (2004b) Elements of a Computational Theory of Location. *Proceedings of the 7th AGILE Conference on Geographic Information Science*, 29 April - 1 May, Heraklion, Greece

KUHN, W. (2005) Geospatial Semantics: Why, of What, and How? In Spaccapietra, S., March, S. and Aberer, K. (Eds.): *Journal on Data Semantics III, Lecture Notes in Computer Science* **3534**, Berlin: Springer Verlag, pp.1-24

KUHN, W. (2007) An Image-Schematic Account of Spatial Categories. In Winter S. *et al.* (Eds.): *COSIT 2007,* LNCS **4736**, pp.152-168, Springer-Verlag

KUIPERS, B. (1978) Modeling Spatial Knowledge. *Cognitive Science* **2**: 129-153

KUIPERS, B. (1982) The "Map in the Head" Metaphor. *Environment and Behaviour* **14**(2): 202-220

KULIK, L., DUCKHAM, M., EGENHOFER, M. (2005) Ontology-driven map generalization. *Journal of Visual Languages and Computing* **16**: 245-267

LARIVE, M., DUPUY, Y., GAILDRAT, V. (2005) Automatic Generation of Urban Zones. *WSCG'2005*, Plzen, Czech Republic, 31 January-4 February, pp.9-12

LEE, D. (2004) Geographic and cartographic contexts in generalisation. *ICA Workshop on Generalisation and Multiple representations*, 20-21 August, Leicester

LÉVI-STRAUSS, C. (1963) *Structural Anthropology*. New York: Basic Books

LEWIS, D. (1970) General semantics. *Synthese* **22**(1-2): 18-67

LEWIS, D. D., SPARCK JONES, K. (1996) Natural language processing for information retrieval. *Communications of the ACM* **39**(1): 92-101

LI, Z., YAN, H., AI, T., CHEN, J. (2004) Automated building generalisation based on urban morphology and Gestalt theory. *International Journal of Geographical Information Science* **18**(5): 513-534

LIEDTKE, C.-E., BÜCKNER, J., GRAU, O., GROWE, S., TÖNJES, R. (1997) AIDA: A system for the knowledge based interpretation of remote sensing data. Third International Airborne Remote Sensing Conference and Exhibition, 7-10 July, Copenhagen, Denmark

LIND, M. (1999) Making sense of the abstraction hierarchy. In: Hoc, J. M., Millot, P., Hollnagel, E. and Cacciabue, P. C. (Eds.): *Proceeding of the Seventh European Conference on Cognitive Science Approaches to Process Control, CSAPC'99*, Villeneuve d'Ascq, France, 21-24 September, pp.195-200

LIU, Y., MOLENAAR, M., AI, T., LIU, Y. (2003) Categorical database generalisation aided by data model. *Proceedings of the 21st International Cartographic Conference (ICC)*, Durban, South Africa, 10-16 August, pp.2308-2318

LLOYD, R., HEIVLY, C. (1987) Systematic Distortions in Urban Cognitive Maps. *Annals of the Association of American Geographers* **77**(2): 191-207

LONGLEY, P. A., HARRIS, R. J. (1999) Towards a new digital data infrastructure for urban analysis and modelling. *Environment and Planning B: Planning and Design* **26**(6): 855-878

LONGLEY, P. A., MESEV, V. (2000a) Measuring Urban Morphology using Remotely-Sensed Imagery. Chapter 9, In: Donnay, J.-P., Barnsley, M. J., Longley, P. A.

(Eds.): *Remote Sensing and Urban Analysis*, Volume 1, Part 4, New York: Routledge, pp.163-184

LONGLEY, P. A., MESEV, V. (2000b) On the measurement and generalisation of urban form. *Environment and Planning A* **32**(3): 473-488

LOPEZ, X., ANNAMALAI, M. (2006) Developing Semantic Web Applications using the Oracle Database 10g RDF Data Model. *Oracle OpenWorld Conference*, 22-26 October, San Francisco, CA, USA, available from URL: http://www.oracle.com/technology/tech/semantic_technologies/pdf/oow2006_semantics_061128.pdf [Last Accessed: 01/12/2008]

LÓPEZ, F. (1999) Overview of methodologies for building ontologies. *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends*, CEUR Publications

LOWE, J. (2008) Make Mashups Correct, Complete, Relevant and Revisited. *Where 2.0 Conference*, 12-14 Mai, London

LOWENTHAL, D. (1961) Geography, experience, and imagination: Towards a geographical epistemology. *Annals of the Association of American Geographers* **51**: 241–60.

LU, W., HAN, J., OOI, B. C. (1993) Discovery of General Knowledge in Large Spatial Databases. *Proc. Far East Workshop on Geographic Information Systems,* Singapore, pp.275-289

LUCK, M., WU, J. (2002) A gradient analysis of urban landscape pattern: a case study from the Phoenix metropolitan region, Arizona, USA. *Landscape Ecology* **17**: 327-339

LUTZ, C., MÖLLER, R. (1997) Defined topological relations in description logics. In Rousset, M.-C. *et al.* (Eds.): *Proceedings of the International Workshop on Description Logics (DL'97),* 27-29 September, Gif sur Yvette, France

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

LÜSCHER, P., BURGHARDT, D., WEIBEL, R. (2007) Ontology-driven enrichment of spatial databases. *10th ICA Workshop on Generalisation and Multiple Representation*, 2-3 August, Moscow

LÜSCHER, P., WEIBEL, R., MACKANESS, W. A. (2008) Where is the terraced house? On the use of ontologies for recognition of urban concepts in cartographic databases. In Ruas, A. and Gold, C. (Eds.): *Headway in Spatial Data Handling, Proceedings of the 13th International Symposium on Spatial Data Handling (SDH)*, 23-25 June, Montpellier, France, Lecture Notes in Geoinformation and Cartography, pp.449-466

LYNCH, K. (1960) *The image of the city*. Cambridge, MA: MIT Press

MACKANESS, W., EDWARDS, G. (2002) The Importance of Modelling Pattern and Structure in Automated Map Generalisation. *Joint Workshop on Multi-Scale Representations of Spatial Data*, 7-8 July, Ottawa, Canada

MANDL, P., HALL, M. M. (2006) HarmonISA − ein Geoinformationswerkzeug zum Studium grenzüberschreitender Landnutzung. In Strobl, J., Blaschke, T. und

Griesebner, G. (Hrsg.): *Angewandte Geoinformatik 2006.* Beiträge zum 18. AGIT-Symposium Salzburg, Herbert Wichman Verlag, Heidelberg

MANDLER, J. (1983) Representation. In Mussen, P. (Ed.): *Handbook of Child Psychology*, Vol. III. New York: John Wiley & Sons, pp. 420-494

MARK, D. M., EGENHOFER, M., HIRTLE, S., SMITH, B. (2004) Ontological Foundations for Geographic Information Science. In McMaster, R. B. and Usery, L. (Eds.): *Research Challenges in Geographic Information Science*, pp. 335-350, Boca Raton, FL: CRC Press

MARK, D. M., FRANK, A. U. (1996) Experiential and formal models of geographic space. *Environment and Planning B: Planning and Design* **23**(1): 3 – 24

MARK, D. M., FREKSA, C., HIRTLE, S. C., LLOYD, R., TVERSKY, B. (1999) Cognitive models of geographical space. *International Journal of Geographical Information Science* **13**(8): 747-774

MARK, D. M., SMITH, B. (2005) A science of topography: Bridging the qualitative-quantitative divide. In *Geographic Information Science and Mountain Geomorphology*. Springer-Praxis

MARLES, A. C. (1983) Identifying and meeting map user needs. *2nd International Symposium on Cartographic Communication*, London

MARR, D. (1982) *Vision.* San Francisco: W. H. Freeman and Company

MARSHALL, S. (2005) *Streets & Patterns*. Abingdon: Spon Press

MATHER, P. M. (1999) Computer Processing of Remotely-Sensed Images, 2nd ed. Chichester: John Wiley & Sons

MASOLO, C., BORGO, S., GANGEMI, A., GUARINO, N., OLTRAMARI, A. (2003) *WonderWeb Deliverable D18*. Laboratory for Applied Ontology, Institute of Cognitive Science and Technology, Italian National Research Council, available from URL: http://www.loa-cnr.it/DOLCE.html [Last Accessed: 27/09/2008]

MATSUYAMA, T., HWANG, V. (1985) SIGMA: A framework for image understanding: Integration of bottom-up and top-down analyses. *Proc. of IJCAI85*, pp.908-915

MAYER, H. (1999) Automatic Object Extraction from Aerial Imagery – A Survey Focusing on Buildings. *Computer Vision and Image Understanding* **74**(2): 138-149

MCCAULEY, S., GOETZ, S. J. (2004) Mapping residential density patterns using multi-temporal Landsat imagery and a decision-tree classifier. International Journal of Remote Sensing 25(6):1077-1094.

MCGUINNES, D. L. AND VAN HARMELEN, F. (2004) OWL Web Ontology Language Overview. World Wide Web Consortium, March 2003. Available from URL: http://www.w3.org/TR/owl-features/ [Last accessed: 21/07/2009]

MENNIS, J. L. (2003) Derivation and implementation of a semantic GIS data model informed by principles of cognition. *Computers, Environment and Urban Systems* **27**: 455-479

MENNIS, J. L., PEUQUET, D. J., QIAN, L. (2000) A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* **14**(6): 501-520

MESEV, V. (2003) Neighbourhood pattern recognition from mailing information: links with satellite imagery. *Journal of Space Communication*, Issue 3

MILLER, D. R. (1994) Knowledge-Based Systems Coupled with Geographic Information Systems. In *Proceedings of the International Workshop on Advanced Information Systems: Geographic Information Systems,* Lecture Notes in Computer Science **884**:143-154, Springer-Verlag

MILLER, G. A. (1978) Practical and Lexical Knowledge. In Rosch, E. and Lloyd, B. B. (Eds.): *Cognition and Categorization*. New Jersey: Lawrence Erlbaum Associates, pp. 305-319

MILLER, G. A. (2003) The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences* **7**(3): 141-144

MILLER, H. J. (2004) Tobler's First Law and Spatial Analysis. *Annals of the Association of American Geographers* **94**(2): 284-289

MINSKY, M. (1975) A Framework for Representing Knowledge. In Winston, P. H. (Ed.): *The Psychology of Computer Vision*, New York: McGraw-Hill, pp.211-277

MIZEN, H., DOLBEAR, C., HART, G. (2005) Ontology Ontogeny: Understanding how an Ontology is created and developed. In Rodríguez, M. A., Cruz, I. F., Egenhofer, M. J. and Levashkin, S. (Eds.): *GeoSpatial Semantics*, First International Conference, GeoS 2005, Mexico City, Mexico, 29-30 November, Lecture Notes in Computer Science **3799**, pp.15-29, Springer Verlag

MONTELLO, D. R. (2002) Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches. *Cartography and Geographic Information Science* **29**(3): 283-304

MONTES DE OCA, V., TORRES, M., LEVACHKINE, S., MORENO, M. (2006) Spatial Data Description by Means of Knowledge-Based System. In Martínez-Trinidata, J. F. *et al.* (Eds.): *CIARP 2006*, LNCS 4225, pp. 502-510, Springer-Verlag

MOTIK, B., HORROCKS, I. (2008) OWL Datatypes: Design and Implementation. In Sheth, A. *et al.* (Eds.): *The Semantic Web – ISWC 2008*, 7th International Semantic Web Conference, Karlsruhe, Germany, 26-30 October, Lecture Notes in Computer Science **5318**, pp. 307-322, Springer-Verlag

MOTIK, B., CUENCA GRAU, B., HORROCKS, I., WU, Z., FOKOUE, A., LUTZ, C. (Eds.) (2008) OWL 2 Web Ontology Language: Profiles. W3C Working Draft 02 December 2008, available from URL: http://www.w3.org/TR/2008/WD-owl2-profiles-20081202/ [Last accessed: 30/01/2009]

MÖLLER, R., HAARSLEV, V. (2008) Tableau-based Reasoning. In Staab, S. and Studer, R. (Eds.): *Handbook of Ontologies*. Springer, to appear

MÖLLER, R., HAARSLEV, V., NEUMANN, B. (2000) Expressive Description Logics for Agent-Based Information Retrieval. In Cuena, J., Demazeau, Y., Garcia, A.,

and Treur, J. (Eds.): *Knowledge Engineering and Agent Technology, Volume 52, Frontiers in Artificial Intelligence and Applications*. Amsterdam: IOS Press

MÖLLER, R., NEUMANN, B., WESSEL, M. (1999) Towards computer vision with description logics: some recent progress. *Proceedings of Integration of Speech and Image Understanding*, Corfu, Greece, pp.101-115

MÖLLER, R., WESSEL, M. (1999) Terminological Default Reasoning About Spatial Information: A First Step. In Freksa, C., and Mark, D. M. (Eds.): *Spatial Information Theory, Cognitive and Computational Foundations of geographic Information Science*, LNCS 1661, International Conference COSIT, Springer-Verlag

MULDER, J. A., MACKWORTH, A. K., HAVENS, W. S. (1988) Knowledge Structuring and Constraint Satisfaction: The Mapsee Approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **10**(6): 866-879

MULLALLY, E.-C., O'DONOGHUE, D. P. (2006) Spatial inference with geometric proportional analogies. *Artificial Intelligence Review* **26**(1-2): 129-140

NEISSER, U. (1976) Cognition and Reality: Principles and Implications of Cognitive Psychology. San Francisco: W. H. Freeman and Company

NEUMANN, B. (2005) Cognitive Vision – Towards Generic Models for Scene Interpretation. *SFB/TR 8 Spatial Cognition Colloquium*, Bremen University, available from URL: http://kogs-www.informatik.uni-hamburg.de/~neumann/talks/GenericModels.pdf [Last Accessed: 02/11/2008]

NEUMANN, B., MÖLLER, R. (2008) On scene Interpretation with Description Logics. *Image and Vision Computing* **26**(1): 82-101

NEUMANN, B., SCHRÖDER, S. (1996) How Useful is Formal Knowledge Representation for Image Interpretation? *Workshop on Conceptual Descriptions from Images*, Cambridge, U.K., 19 April, pp.58-69

NEUMANN, B., WEISS, T. (2003) Navigating through logic-based scene models for high-level scene interpretations. *Proceedings of the 3rd International Conference on Computer Vision Systems (ICVS-2003)*, Lecture Notes in Computer Science **2626**, pp. 212-222, Springer-Verlag

NEUN, M., BURGHARDT, D., WEIBEL, R. (2006) Spatial Structures as Generalisation Support Services. *Proceedings of ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data*, 22-24 February, Hannover, Germany, pp.6-15

NEUN, M., STEINIGER, S. (2005) Modelling Cartographic Relations for Categorical Maps. *XXII International Cartographic Conference, ICC2005*, A Coruña, Spain, 11-16 July

NEUN, M., WEIBEL, R., BURGHARDT, D. (2004) Data enrichment for adaptive generalisation. *ICA Workshop on Generalisation and Multiple Representation*, 20-21 August, Leicester

NILES, I., PEASE, A. (2001) Towards a Standard Upper Ontology. *FOIS'01*, 17-19 October, Ogunquit, Maine, USA

NORMAN, D. A. (1999) Affordance, Conventions, and Design. *Interactions* **6**(3): 38-43

NOY, N. F., MCGUINNESS, D. L. (2001) Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory, Available from URL: ftp://ftp.ksl.stanford.edu/pub/KSL_Reports/KSL-01-05.pdf.gz [Last Accessed 07/10/2008]

NUNES, J. (1991) Geographic space as a set of concrete geographical entities. In Mark, D. and Frank, A. (Eds.): *Cognitive and Linguistic Aspects of Geographic Space*, Norwell, MA: Kluwer Academic, pp.9-33

NYULAS, C., O'CONNOR, M., TU, S. (2007) DataMaster – a Plug-in for Importing Schemas and Data from Relational Databases into Protégé. *10th International Protégé Conference*, Budapest, Hungary, 15-18 July

ODPM (2004) *The English Indices of Deprivation 2004*. Neighbourhood Renewal Unit, available from URL: http://www.neighbourhood.gov.uk/page.asp?id=1128440 (last accessed: 04/08/2006)

ODPM (2005) *Generalised Land Use Database Statistics for England*. London: Office of the Deputy Prime Minister, © Crown Copyright, available online from URL: http://www.communities.gov.uk/index.asp?id=1146084 (last accessed: 28/07/2006)

ODPM (2006) *National Land Use Database: Land Use and Land Cover Classification.* Available online from the Department for Communities and Local Government: http://www.communities.gov.uk/index.asp?id=1163845 [Last accessed: 28/07/06], London, ©Queen's Printer and Controller of Her Majesty's Stationery Office 2006.

OPENSHAW, S., OPENSHAW, C. (1997) Artificial Intelligence in Geography. Chichester: John Wiley & Sons

ORDNANCE SURVEY (2002) Ordnance Survey MasterMap: National Land Use Theme Research (Land Use and Land Cover). NOP World, final report, Ordnance Survey, Southampton

ORDNANCE SURVEY (2005-2007) Future Users Research: User-task Interview Records 1-56. Ordnance Survey internal research records, Southampton

ORDNANCE SURVEY (2006) OS MasterMap Address Layer and Address Layer 2 User Guide. Ordnance Survey, Southampton, © Crown copyright

ORLEANS, P. (1973) Differential cognition of urban residents: effects of social scale on mapping. In Downs, R. M. and Stea, D. (Eds.): *Image and Environment*. Chicago: Aldine, pp.115-130

PALMER, S. E. (1978) Fundamental aspects of cognitive representation. In Rosch, E. and Lloyd, B. (Eds.): *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum

PATTON, M. Q. (1990). *Qualitative evaluation and research method*, 2nd ed., Newbury Park, CA: Sage

PAZZANI, M. J. (2000) Knowledge discovery from data? *IEEE Intelligent Systems and Their Applications* **15**(2): 10-12

PEACHAVANISH, R., KARIMI, H. A. (2007) Ontological Engineering for Interpreting Geospatial Queries. *Transactions in GIS* **11**(1):115-130

PECKHAM, J., MARYANSKI, F. (1988) Semantic Data Models. *ACM Computing Surveys* **20**(3): 153-189

PEDERSON, T., PATWARDHAN, S., MICHELIZZI, J. (2004) WordNet::Similarity – Measuring the Relatedness of Concepts. In Proceedings of the 19th National Conference on Artificial Intelligence (AAAI-04), 25-29 July, San Jose, CA, pp. 1024-1025

PERDIKOGIANNI, I. (2003) Heraklion and Chania: A study of the evolution of their spatial and functional patterns, *Proceedings of the 4th International Space Syntax Symposium* London

PESARESI, M., BIANCHIN, A. (2000) Recognizing settlement structure using mathematical morphology and image texture. Remote Sensing and Urban Analysis, GISDATA 9, pp. 55-67

PETER, B. (2001) Measures for the Generalization of Polygonal Maps with Categorical Data. *Fourth ICA Workshop on Progress in Automated Map Generalization*, Beijing, 2-4 August

PEUQUET, D.J. (1988) Representations of Geographic Space: Toward a Conceptual Synthesis. *Annals of the Association of American Geographers* **78**(3): 375-394

PEUQUET, D., SMITH, B., BROGAARD, B. (1998) The Ontology of Fields. Technical report, NCGIA, University of California at Santa Barbara, Report of a specialist meeting held under the auspices of the Varenius Project, Bar Harbour, Maine, 11-13 June

PINCHEMEL, P. (1995) La face de la terre eléments de géographie. Armand Colin

PINKER, S. (1991) Rules of Language. *Science* **253**(5019): 530-535

PINKER, S. (1997) *How the mind works.* London: Allen Lane The Penguin Press

POGGI, A., LEMBO, D., CALVANESE, D., DE GIACOMO, G., LENZERINI, M., ROSATI, R. (2008) Linking Data to Ontologies. *Journal on Data Semantics* **X**:133-173

POL, P. M. J. (2002) The necessity of analysing cities in a comprehensive way. *42nd European RSA Congress*, Dortmund

POLI, R. (1996) Ontology for Knowledge Organisation. In Green, R. (Ed.): *Knowledge organization and change: Proceedings of the 4th international ISKO conference*, 15-18 July, Washington D.C., Frankfurt: Indeks-Verlag, pp.313-319

POPE, C., MAYS, N. (1995) Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. *BMJ* **311**: 42-45

POPE, C., ZIEBLAND, S., MAYS, N. (2000) Analysing qualitative data. *BMJ* **320**:114-116

RAM, S. (1995) Intelligent database design using the unifying semantic model. *Information & Management* **29**: 191-206

RAPER, J., LIVINGSTONE, D. (1995) Development of a geomorphological spatial model using object-oriented design. *International Journal of Geographical Information Systems* **9**(4): 359-383

RECTOR, A. (2004) Why use a Classifier? When will it help? And when will it not? *7th International Protégé Conference*, Bethesda, Maryland, 6-9 July

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

REGNAULD, N. (2006) Open architecture to provide on demand mapping. Ordnance Survey Research, internal report, Southampton, ©Crown Copyright

REGNAULD, N. (2008) Developing generalisation systems with reusability in mind. Ordnance Survey Research, internal report, Southampton, ©Crown Copyright

REITER, R., MACKWORTH, A. K. (1989) A logical framework for depiction and image interpretation. *Artificial Intelligence* **41**: 125-155

RICHARDSON, V. (1996) The role of attitudes and beliefs in learning to teach. In Sikula, J. (Ed.): *The handbook of research on teacher education*, 2nd ed., pp.102-119 New York: Macmillan

RIPS, L. J. (1994) *The Psychology of Proof: Deductive Reasoning in Human Thinking.* Cambridge, MA: MIT Press

RODRÍGUEZ, M. A., GODOY, F. A. (2002) A content-based approach to searching and indexing spatial configurations. In Egenhofer, M.J. and Mark, D.M. (Eds.): *GIScience'02,* Germany, Lecture Notes in Computer Science, Vol. 2478, Springer-Verlag, pp.260-275

RODRÍGUEZ, M. A., JARUR, M. C. (2005) A genetic algorithm for searching spatial configurations. *IEEE Transactions on Evolutionary Computation* **9**(3): 252-270

RODRÍGUEZ, M. A., EGENHOFER, M. J., RUGG, R. D. (1999) Assessing semantic similarities among geospatial feature class definitions. In Vckovski, A., Brassel, K. and Schek, H.-J. (Eds.): *Interoperating Geographic Information Systems,* Interop'99, Zurich, Switzerland, LNCS 1580: 189-202, Springer-Verlag

ROSCH, E. (1978) Principles of Categorization. In Rosch, E. and Lloyd, B.B. (Eds.): *Cognition and Categorization.* New Jersey: Lawrence Erlbaum Associates

ROTHKEGEL, R., WENDER, K.F., SCHUMACHER, S. (1998) Judging Spatial Relations from Memory. In Freksa, C., Habel, C. and Wender, K.F. (Eds.): *Spatial Cognition An Interdisciplinary Approach to Representing and Processing Spatial Knowledge, Lecture Notes in Artificial Intelligence* **1404**: 79-105 Berlin: Springer Verlag

RUAS, A. (2000) The roles of meso objects for generalisation. *Proceedings of the 9th International Symposium on Spatial Data Handling,* 10-12 August, Beijing, China, pp.50-63

RUGG, R., EGENHOFER, M., KUHN, W. (1997) Formalizing Behavior of Geographic Feature Types. *Geographical Systems* **4**(2): 159-180

RUSSELL, S. J., NORVIG, P. (1995) *Artificial Intelligence: A Modern Approach.* New Jersey: Prentice Hall

SANTOS, P., STORK, A., LINAZA, M. T., MACHUI, O., MCINTYRE, D., JORGE, E. (2007) CINeSPACE: Interactive Access to Cultural Heritage While On-The-Move. In Schuler, D. (Ed.): *Online Communities and Social Computation, HCII2007, LNCS* **4564**: 435-444, Berlin: Springer-Verlag

SCHAERF, A. (1994) Reasoning with Individuals in Concept Languages. *Data and Knowledge Engineering* **13**(2): 141-176

SCHRÖDER, C. (1999) *Bildinterpretation durch Modellkonstruktion: Eine Theorie zur rechnergestützten Analyse von Bildern.* Sankt Augustin: Infix

SCHWERING, A. (2005) Hybrid model for semantic similarity measurement. In Meersman, R. and Tari, Z. (Eds.): *CoopIS/DOA/ODBASE 2005*, LNCS 3761, pp.1449-1465, Springer-Verlag

SCHWERING, A., RAUBAL, M. (2005) Measuring semantic similarity between geospatial conceptual regions. In Rodríguez *et al.* (Eds.): GeoS 2005, LNCS 3799, pp.90-106, Springer-Verlag

SCHWERING, A., WAGNER, R. M., SCHNEIDER, B. (2003) A geographic search: Ways to structure information on the web. In *Proceedings of the 6th AGILE Conference on GIScience*, Lyons, France, 24-26 April

SCHWITTER, R., TILBROOK, M. (2006) Let's Talk in Description Logic via Controlled Natural Language. *Proceedings of the Third International Workshop on Logic and Engineering of Natural Language Semantics (LENLS2006)*, 20th Annual Conference of the Japanese Society for Artificial Intelligence, Tokyo, Japan, 5-6 June, pp.193-207

SCHWITTER, R., KALJURAND K., CREGAN, A., DOLBEAR, C., HART, G. (2008) A Comparison of three Controlled Natural Languages for OWL 1.1. *4th OWL Experiences and Directions Workshop, OWLED 2008*, Washington DC, 1-2 April

SESTER, M. (2000) Knowledge acquisition for the automatic interpretation of spatial data. *International Journal of Geographical Information Science* **14**(1): 1-24

SIRIN, E., PARSIA, B., CUENCA GRAU, B., KALYANPUR, A., KATZ, Y. (2007) Pellet: A Practical OWL-DL Reasoner. *Journal of Web Semantics* **5**(2): 51-53

SMITH, B. (2004) Beyond Concepts: Ontology as Reality Representation. In Varzi, A. and Vieu, L. (Eds.): *Proceedings of the International Conference on Formal Ontology and Information Systems FOIS 2004*, Turin, 4-6 November

SMITH, B., CEUSTER, W., TEMMERMAN, R. (2005) Wüsteria. In Engelbrecht, R., Geissbuhler, A., Lovis, C. and Mihalas, G. (Eds.): *Connecting Medical Informatics and Bio-Informatics. Proceedings of MIE2005 – The XIXth International Congress of the European Federation for Medical Informatics* Vol. 116, pp.647-652

SMITH, B., MARK, D.M. (1998) Ontology and Geographic Kinds. In *Proceedings of the International Symposium on Spatial Data Handling (SDH'98)*, 12-15 July Vancouver, Canada

SMITH, B., MARK, D. M. (2001) Geographical categories: an ontological investigation. *International Journal of Geographical Information Science* **15**(7): 591-612

SMITH, B., VARZI, A. C. (2000) Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological Research* **60**(2): 401–420

SOWA, J. F. (2000) *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Pacific Grove, CA: Brooks Cole Publishing

SOWMYA, A., TRINDER, J. (2000) Modelling and representation issues in automated feature extraction from aerial and satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing* **55**: 34-47

SPERBER, D., HIRSCHFELD, L. (1999) Culture, Cognition, and Evolution. In Wilson, R.A. and Keil, F.C. (Eds.): *The MIT Encyclopedia of the Cognitive Sciences*, pp.xv-xxxvii, Massachusetts: The MIT Press

SPYNS, P., MEERSMAN, R., JARRAR, M. (2002) Data modelling versus Ontology engineering. *SIGMOND Record* **31**(4): 12-17

SRIVASTAVA, S. M. (2008) *A Course on Mathematical Logic*. New York: Springer Science+Business Media

STEIMANN, F., NEJDL, W. (1999) Modellierung und Ontologie. Technischer Bericht, Universität Hannover, Institut für Rechnergestützte Wissensverarbeitung, Available from URL [Last Accessed 14/10/2008]: http://www.kbs.uni-hannover.de/Arbeiten/Publikationen/1999/M&O.pdf

STEINIGER, S. (2006) Classifying urban structures for mapping purposes using discriminant analysis. In Priestnall, G. and Aplin, P. (Eds.): *Proceedings of GISRUK Conference 2006*, Nottingham, 5-7 April, pp.107-111

STEINIGER, S., LANGE, T., WEIBEL, R., BURGHARDT, D. (2008) An Approach for the Classification of Urban Building Structures based on Discriminant Analysis Techniques. *Transactions in GIS* **12**(1): 31-59

STEINIGER, S., WEIBEL, R. (2005) Relations and Structures in Categorical Maps. *8th ICA Workshop on Generalisation and Multiple Representation*, A Coruña, Spain, 7-8 July

STENMARK, D. (2001) The Relationship between Information and Knowledge, in *Proceedings of IRIS 24*, Ulvik, Norway, August 11-14

STEVENS, R., EGAÑA ARANGUREN, M., WOLSTENCROFT, K., SATTLER, U., DRUMMOND, N., HORRIDGE, M., RECTOR, A. (2007) Using OWL to Model Biological Knowledge. *International Journal of Human-Computer Studies* **65**(7): 583-594

STUCKENSCHMIDT, H., VAN HARMELEN, F. (2005) *Information Sharing on the Semantic Web*. Berlin: Springer Verlag

SUI, D. Z. (1996) Contextualizing Geographic Information Systems (GIS): Toward a critical theory of geographic information science. *GIS and Society: The social implications of how people, space, and environment are represented in GIS – Proceedings of NCGIA Initiative 19 Specialist Meeting*, pp.30-35

SVÁTEK, V. (2006) Ontologies, questionnaires and (mining) tabular data. In *the 3rd European Semantic Web Conference ESWC 2006*, Budva, Montenegro, 11-14 June

SVÁTEK, V., RAUCH, J., RALBOVSKÝ, M. (2006) Ontology-Enhanced Association Mining. In Ackermann *et al*. (Eds.): *EWMF/KDO 2005*, LNAI **4289**, pp.163-179, Springer-Verlag

TANASESCU, V. (2007) Spatial Semantics in Difference Spaces. In Winter, S., Duckham, M., Kulik, L., Kuipers, B. (Eds.): *Spatial Information Theory, 8th*

International Conference, COSIT 2007, Melbourne, Australia, 19-23 September, Lecture Notes in Computer Science 4736, Springer Verlag pp.96-115

TANG, A. Y., ADAMS, T. M., USERY, E. L. (1996) A spatial data model design for feature-based geographical information systems. *International Journal of Geographical Information Systems* **10**(5): 643-659

THOMSON, M.-K. (2007) Questionnaires for Ontology Engineering. In Probst, F. and Kessler, C. (Eds.): *GI-Days 2007 – Young Researchers Forum, Proceedings of the 5th Geographic Information Days*, 10-12 September, Munster, Germany, pp. 299-302

THOMSON, M.-K., BÉRA, R. (2007a) Relating Land Use to the Landscape Character: Toward an Ontological Inference Tool. In Winstanley, A. C. (Ed.): *Proceedings of GIS Research UK Conference*, *GISRUK'07*, Maynooth, Ireland, 11-13 April 2007, pp. 83-87

THOMSON, M.-K., BÉRA, R. (2007b) Relating Land Use to the Landscape Character: Toward an Ontological Inference Tool. *SAGEO 2007*, 18-19 June, Clermont-Ferrand, France

THOMSON, M.-K., BÉRA, R. (2008) A methodology for inferring higher level semantic information from spatial databases. In Lambrick, D. (Ed.): Proceedings of the GIS Research UK 16th Annual Conference, GISRUK'08, UNIGIS, Manchester, 2-4 April, pp.268-274

TIMPF, S., FRANK, A. U. (1997) Using hierarchical spatial data structures for hierarchical spatial reasoning. In Hirtle, S. C. and Frank, A. U. (Eds.): *Spatial Information Theory – A Theoretical Basis for GIS*, International Conference COSIT'97, Lecture Notes in Computer Science **1329**: 69-83, Springer-Verlag

TOBLER, W. R. (1970). "A computer movie simulating urban growth in the Detroit region". *Economic Geography* **46**(2): 234-240.

TOLMAN, E. C. (1948) Cognitive maps in rats and men. *Psychological Review* **55**: 189-208

TOMPKINSON, W., MORTON, D., GOMM, S., SEAMAN, E. (2004) An optimised semi-automated methodology for populating a national land use dataset. *Proceedings of the 12th Annual Geographical Information Science Research UK Conference (GISRUK '04)*, University of East Anglia, Norwich, 28-30 April, pp.35 – 38

TORRES, M., QUINTERO, R., MORENO, M., FONSECA, F. (2005) Ontology-driven description of spatial data for their semantic processing. In Rodríguez *et al.* (Eds.): *GeoS 2005*, *LNCS 3799*, Springer-Verlag, pp.242-249

TORRES, M., GUZMÁN, G., QUINTERO, R., MORENO, M., LEVACHKINE, S. (2006) Semantic Decomposition of LandSat TM Image. In Gabrys, B., Howlett, R. J. and Jain, L. C. (Eds.): *KES 2006, Part I*, LNAI **4251**, pp.550-558, Springer-Verlag

TROWBRIDGE, C. C. (1913) On fundamental methods of orientation and imaginary maps. *Science* **38**: 990

TVERSKY, B. (1993) Cognitive maps, cognitive collages, and spatial mental models. In Frank, A. U. and Campari, I. (Eds.): *Spatial Information Theory: A*

Theoretical Basis for GIS, Proceedings COSIT'93. Lecture Notes in Computer Science, Vol. 716, pp.14-24, Berlin: Springer-Verlag

TVERSKY, B., LEE, P. U. (1998) How Space Structures Language. In Freksa, C., Habel, C. and Wender, K. F. (Eds.): *Spatial Cognition An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*, Lecture Notes in Artificial Intelligence, Vo. 1404, pp.157-176 Berlin: Springer-Verlag

USCHOLD, M., GRÜNINGER, M. (2004) Ontologies and Semantics for Seamless Connectivity. *SIGMOD Record* **33**(4): 58-64

VALLET, D., FERNÁNDEZ, M., CASTELLS, P. (2005) An Ontology-Based Information Retrieval Model. *Proceedings of the 2ⁿᵈ European Semantic Web Conference, ESWC 2005*, LNCS **3532**, pp.455-470

VAUGHAN, L., CHATFORD CLARK, D. L., SAHBAZ, O., HAKLAY, M. (2005) Space and Exclusion: Does urban morphology play a part in social deprivation? *Area* **37**(4): 402-412

VERASTEGUI, K., MARTINEZ, M., MORENO, M., LEVACHKINE, S., TORRES, M. (2006) Incorporating semantics into GIS applications. In Gabrys, B., Howlett, R. J. and Jain, L. C. (Eds.): *KES 2006, Part I*, LNAI 4251, pp.698-705, Springer-Verlag

VIEU, L. (1993) A Logical Framework for Reasoning about Space. In Frank, A.U. and Campari, I. (Eds.): *Spatial Information Theory. A Theoretical Basis for GIS.* Proceedings of the European Conference on Spatial Information Theory (COSIT'93), Elba, Italy, pp.25-35

VILLANUEVA-ROSALES, N., OSBAHR, K., DUMONTIER, M. (2007) Towards a semantic knowledge base for yeast biologists. *Proceedings of the First International Workshop on Health Care and Life Sciences Data Integration for the Semantic Web (HCLS 2007)*, Banff, Canada

VISSER, U., STUCKENSCHMIDT, H., SCHUSTER, G., VÖGELE, T. (2000) Ontologies for Geographic Information Processing. *Computers & Geosciences* **28**(1): 103-118

VORWERG, C., RICKHEIT, G. (1998) Typicality Effects in the Categorization of Spatial Relations. In Freksa, C., Habel, C. and Wender, K.F. (Eds.): *Spatial Cognition An Interdisciplinary Approach to Representing and Processing Spatial Knowledge. Lecture Notes in Artificial Intelligence* **1404**: 203-222 Berlin: Springer-Verlag

WAHL, B. (2008) Mapping the World … One Neighborhood at a Time. *Directions Magazine*, article from 1ˢᵗ and 3ʳᵈ of December, available from URL: http://www.directionsmag.com/article.php?article_id=2944&trv=1   [Last Accessed: 01/12/2008]

WALL, C. R., DEHAVEN, M. J., OEFFINGER, K. C. (2002) Survey methodology for the uninitiated. *The Journal of Family Practice* **51**(6):1-11

WANG, H. H., NOY, N., RECTOR, A., MUSEN, M., REDMOND, T., RUBIN, D., TU, S., TUDORACHE, T., DRUMMOND, N., HORRIDGE, M., SEIDENBERG, J. (2006) Frames and OWL Side by Side. *9ᵗʰ International Protégé Conference*, Stanford, USA

WATERS, T., EVANS, A. J. (2003) Tools for web-based GIS mapping of a "fuzzy" vernacular geography. *7th International Conference on GeoComputation,* Southampton, 8-10 September

WEIBEL, R., KELLER, S., REICHENBACHER, T. (1995) Overcoming the knowledge acquisition bottleneck in map generalization: The role of interactive systems and computational intelligence. *COSIT'95, Lecture Notes in Computer Science*, **988**: 139-156 Springer-Verlag

WERTHEIMER, M. (1924) Gestalt Theory. An address before the Kant Society, Berlin, 7th December, 1924. In the translation by Ellis, W. D. (1938) *Source Book of Gestalt Psychology*. New York: Harcourt, Brace and Co

WESSEL, M., MÖLLER, R. (2006) A Flexible DL-based Architecture for Deductive Information Systems. *Proceedings of the IJCAR-06 Workshop on Empirically Successful Computerized Reasoning (ESCOR)*, pp.92-111

WESSEL, M., MÖLLER, R. (2007) Flexible software architectures for ontology-based information systems. *Journal of Applied Logic, Special Issue on Empirically Successful Systems*, in press

WIEGAND, N., PATTERSON, D., ZHOU, N., VENTURA, S., CRUZ, I. F. (2002) Querying heterogeneous land use data: Problems and potential. In *National Conference for Digital Government Research*, pp115-121

WIELINGA, B., SCHREIBER, G. (1997) Configuration Design Problem Solving. *IEEE Expert/Intelligent Systems and their Applications* **12**(2): 49-56

WILSON, R. A., KEIL, F. C. (Eds.) (1999) The MIT Encyclopedia of the Cognitive Sciences. Massachusetts: The MIT Press

WLODARCZYK, D. (2005) *Structural Analysis of urban space in residential areas. In Methodologies in Housing Research*, Newcastle upon Tyne: The Urban International Press

WOLSTENCROFT, K., LORD, P., TABERNERO, L., BRASS, A., STEVENS, R. (2006) Protein classification using ontology classification. *Bioinformatics* **22**(14): e530-e538

WOLTER, F., ZAKHARYASCHEV, M. (2005) A logic for metric and topology. *Journal of Symbolic Logic* **70**: 795-828

WORBOYS, M. (1995) *GIS – A computing perspective*. London: Taylor and Francis

WORBOYS, M. (1996) Metrics and topologies for geographic space. In Kraak, M. J. and Molenaar, M. (Eds.): *Advances in GIS Research II*. London: Taylor and Francis, pp. 365-375

WYATT, P. (2002) Creation of an urban land use database, The Education Trust of The Royal Institution of Chartered Surveyors, Available from URL: http://www.rics-educationtrust.org/publish/download.aspx?did=3069 (Last Accessed: 05/11/2005)

WYATT, P. (2004) Constructing a Land-Use Data Set from Public Domain Information in England, *Planning, Practice & Research* **19**(2): 147-171

ZEGARELLI, M. (2007) *Logic for Dummies*. Hoboken, NJ: Wiley Publishing

ZHANG, Q. (2004) Modelling structure and patterns in road network generalisation. *ICA Workshop on Generalisation and Multiple Representation*, 20-21 August, Leicester, UK.

ZHANG, Y.-J. (Ed.) (2007) *Semantic-Based Visual Information Retrieval.* Hershey: IRM Press

ZHAO, T., ZHANG, C., WEI, M., PENG, Z.-R. (2008) Ontology-based geospatial data query and integration. In Cova, T. J., Miller, H. J., Beard, K., Frank, A. U. and Goodchild, M. F. (Eds.): *Geographic Information Science, 5th International Conference GIScience 2008*, LNCS **5266**, Park City, UT, USA, 23-26 September, Berlin: Springer-Verlag

ZLATOFF, N., TELLEZ, B., BASKURT, A. (2004) Image understanding and scene models: a generic framework integrating domain knowledge and gestalt theory. *In Proceedings of International Conference on Image Processing, ICIP'04,* Vol.4, pp.2355-2358

Appendix A

Questionnaire

Attached is the questionnaire that was used as a tool to elicit knowledge about the interpretation process of topographic maps as well as the conceptualisation of the land use domain, as described in chapter 3.

Relating Land Use to the Landscape Character

A study about human knowledge, map interpretation and reasoning skills. By Tina Thomson, PhD student at the Department of Geomatic Engineering, UCL, London.

INFORMATION PAGE

Introduction
Your participation is greatly appreciated and will contribute towards an Engineering and Physical Sciences Research Council (EPSRC) and Ordnance Survey sponsored project, which aims to improve geographic data.

I would like your help with this research because I am trying to learn how people understand and interpret different types of land use. The term land use refers to the way humans make use of a piece of land.

Procedure
- The questionnaire will take up to an hour to complete.
- You will need to have a few coloured pens or pencils to hand before you start.
- Please follow through the questionnaire from start to end in the order it has been laid out.
- Please don't use any other resources while completing it.
- Instructions for completing each question are included in brackets.
- Please try to write as clearly as possible.

Confidentiality Statement
By taking part in this research your answers will remain confidential. You will be asked for some personal information at the end of this exercise in order to evaluate your responses more effectively. However, you will not be asked for any information that identifies you individually.

I hope you find this questionnaire interesting and fun. For any further questions or help, please do not hesitate to contact me at thomson@ge.ucl.ac.uk

Thanks again,
Tina.

Task 1: Map Interpretation

On the next two pages you will find two plain topographic maps with no additional cartographic information, colouring, scale or orientation information. Only the roads have been coloured in grey.

Your task is to carefully examine the maps. Try to look for features you may recognise, similarities and patterns. Then group those objects that you believe belong to the same category by circling or colouring their area with a highlighter or different coloured pen.

Map A
For the first map imagine the following scenario: You are buying a house in the shown area. You are interested where for example the nearest shops, school, GP, parks, etc. are located. You want to know what you can **use** the areas for.

Now try to interpret this map according to how one can **use** the areas by colouring them in. Give each '**use**' you identify a name that you find most appropriate for describing it. Write the names in the column linking it to the area they refer to (same colour, or drawing a line to the area), and state how confident you are about your decision (1 - very confident, 2 - somewhat confident, 3 - not confident).

Map B
For the second map imagine the following scenario: urban planning is concerned with improving the landscape. Areas are designed and organised in an aesthetic and effective way. Everything is built for a specific **purpose**.

Now try to interpret this map according to the **purpose** of areas by colouring them. Give each **purpose** you identify a name that you find most appropriate for describing it. Again, write the names in the column linking it to the area they refer to (same colour, or drawing a line to the area), and state how confident you are about your decision (1 - very confident, 2 - somewhat confident, 3 - not confident).

Please, try to find a minimum of 4 in each map.

ID:_____

Date:____/____/_____
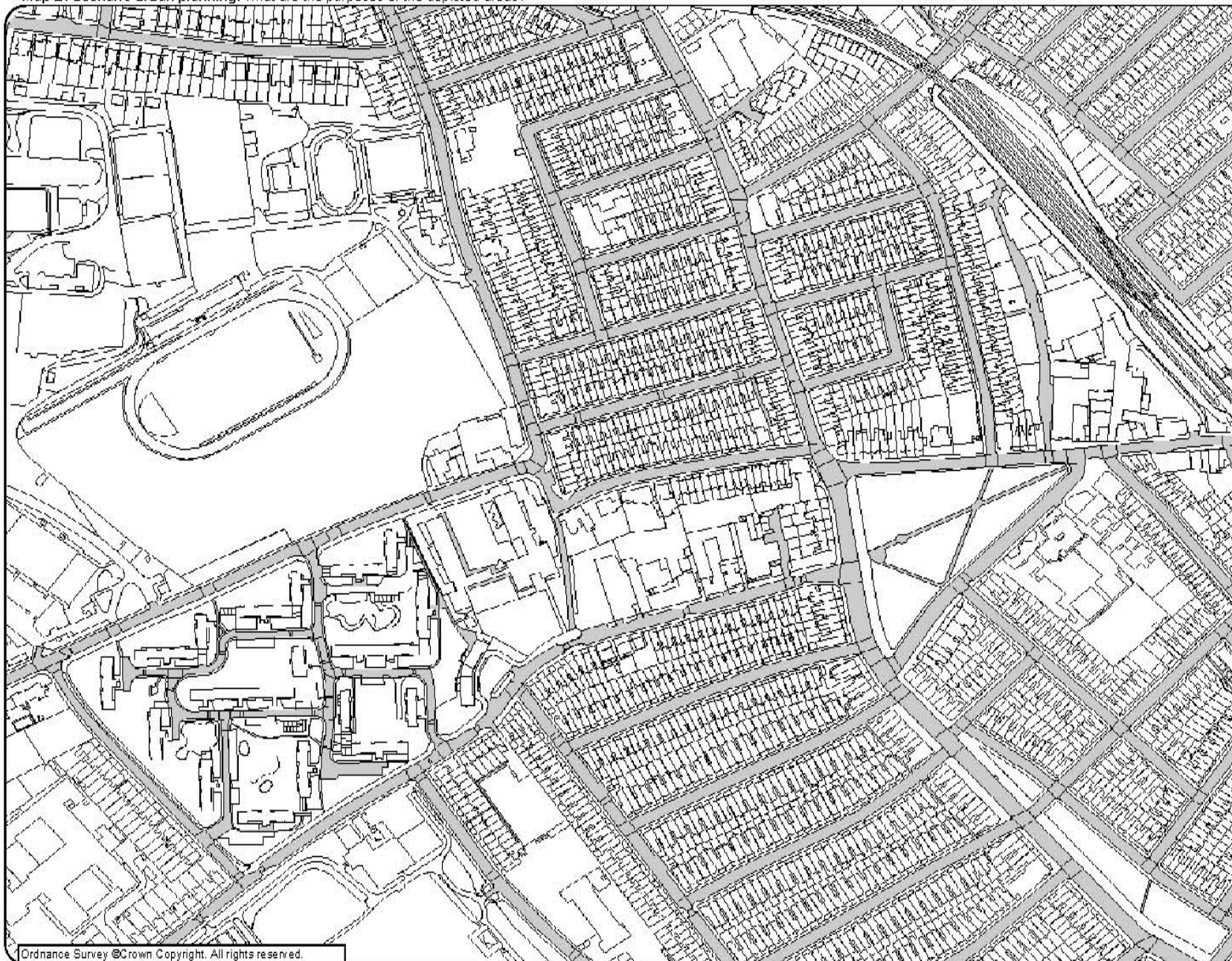
Task 2: Understanding the Interpretation Process

Q.1 How did you approach the map interpretation task? Please describe below.

_____

_____

_____

Q.2 What captured your attention first in:

a) Map A: _____

b) Map B: _____

Q.3 What do you think is the most dominant object throughout each map (e.g. buildings)? And as what did you interpret it (e.g. houses)? Please state below:

	Dominant object:	Interpreted as:
a) Map A:	_____	_____
b) Map B:	_____	_____

Q.4 Did you use areas you already interpreted to identify what the surrounding areas were?
☐Yes ☐No

Q.5 Do you believe there is a repeating pattern for each land use type?
☐Yes ☐No

Q.6 Would it have helped you if the map showed a bigger area?
☐Yes ☐No

Q.7 a) Map A and map B were not scaled the same. Did the varying scale of the two maps influence your interpretation?
☐Yes ☐No ☐Don't know

b) If yes, in what way? Please describe: _____

_____

Q.8 Was there anything in the map that you weren't sure about? Please explain briefly.

_____

_____

Q.9 a) What did you find difficult during your interpretation? Please describe:

_____

_____

More specifically, do you believe difficulties in the interpretation are caused by any of the following? (Please tick)

	Yes	No
b) Fuzziness of where one land use ends and the other starts	☐	☐
c) Misinterpretation of cartographic objects	☐	☐
d) Not being able to identify an objects' meaning	☐	☐
e) Not being able to identify one area's meaning in relation ot other areas	☐	☐

Q.10 a) Did you recognise any of the locations shown in the maps? ☐Yes ☐No

b) If yes, where do you think the areas are located?

_____

_____

Q.11 Which pieces of information (in relation to other objects) do you think are superior to others in the interpretation process? As far as you are aware, rate the aspects below according to their importance in your interpretation process.

	Not important 1	2	3	4	5 Very important
a) Position	☐	☐	☐	☐	☐
b) Proximity (distance)	☐	☐	☐	☐	☐
c) Shape	☐	☐	☐	☐	☐
d) Size	☐	☐	☐	☐	☐
e) Similarity in arrangement	☐	☐	☐	☐	☐
f) Similarity in geometry	☐	☐	☐	☐	☐
g) Contrast	☐	☐	☐	☐	☐
h) Symmetry	☐	☐	☐	☐	☐
i) Context	☐	☐	☐	☐	☐
j) Simplicity of identification	☐	☐	☐	☐	☐
k) Orientation	☐	☐	☐	☐	☐
l) Connectedness	☐	☐	☐	☐	☐
m) Structure	☐	☐	☐	☐	☐
n) Closure (bound, limit)	☐	☐	☐	☐	☐
o) Common region	☐	☐	☐	☐	☐
p) Good continuation of objects	☐	☐	☐	☐	☐
q) Likelihood of correct interpretation	☐	☐	☐	☐	☐
r) Organisation	☐	☐	☐	☐	☐
s) Topology (e.g. containment, adjacency relations)	☐	☐	☐	☐	☐
t) Combination of objects	☐	☐	☐	☐	☐

Q.12 How important do you rate the following principles for grouping objects together during your interpretation?

	Not important 1	2	3	4	5 Very important
a) Similarity in shape	☐	☐	☐	☐	☐
b) Similarity in size	☐	☐	☐	☐	☐
c) Similarity in orientation	☐	☐	☐	☐	☐
d) The proximity between objects	☐	☐	☐	☐	☐
e) Symmetry in the arrangement	☐	☐	☐	☐	☐
f) Alignment of objects (e.g. horizontal)	☐	☐	☐	☐	☐
g) The overall shape of the group	☐	☐	☐	☐	☐
h) The relation among parts and wholes	☐	☐	☐	☐	☐
i) The influence of one dominating feature within the group	☐	☐	☐	☐	☐

Q.13 Did you apply any other grouping principles for your interpretation? If yes, please list them below.

_____

Q.14 How important do you rate the following for your ability to interpret the maps?

	Not important 1	2	3	4	5 Very important
a) Existing knowledge about land use	☐	☐	☐	☐	☐
b) Memory	☐	☐	☐	☐	☐
c) Experience	☐	☐	☐	☐	☐
d) Awareness of our everyday surroundings	☐	☐	☐	☐	☐
e) Knowledge about the form and sizes of objects constituting a land use	☐	☐	☐	☐	☐
f) Knowledge of what belongs to a land use category	☐	☐	☐	☐	☐

Task 3: Conceptualising a Land Use Type

Choose one of the following land use types (please circle): school, hospital, train station, park, residential, industrial, commercial, outdoor recreation, terraced housing, semi-detached housing, detached housing, public spaces. What do you think constitutes a land use in the landscape?

For your chosen land use try to answer the questions below by entering only one item per box. In the horizontal direction separate categories will define your chosen land use spatially, while in the vertical direction questions will describe each member category in more detail.

Q.15 Member Categories
Imagine yourself in the land use you have chosen. What geographical objects make up the land use? State one per box. Please be specific in terms of what the object is used for. (e.g. car park, garden, etc.)

Q.16 Member Objects
How do they physically look like in the real world? (e.g. building, open space)

Q.17 Purposes
What is their primary purpose?

Q.18 Roles
What are they used for?

Q.19 Affordances
How do people make use of them?

Q.20 Properties
What are their geographical properties in terms of geometry, shape, relative size? (e.g. large, small, square, etc.)

Q.21 Relations
a) Is the member category a 'kind of' (e.g. school is a kind of education) or 'part of' (e.g. play ground is a part of school) of your chosen land use? Please use only one of the two answers 'kind of' or 'part of'.

b) What is the spatial combination of member categories? Use the relations disjoint, meet, overlap, covers, covered by, contains, equal, inside, and state to which one you are relating it to.

c) What is the spatial distance among member categories? Please state if zero, very close, close, or far, and to which one you are relating it to.

Task 4: Personal Information

Personal Data

Q.22 Age group (please tick appropriately): 18-25 ☐ 26-30 ☐ 31-35 ☐ 36-40 ☐ 41+ ☐

Q.23 Gender: Male ☐ Female ☐

Q.24 Nationality: _____

Q.25 Languages spoken (native first): _____

Q.26 a) Where do you live? City (inc. its suburbs) ☐ Town/Village ☐ Rural ☐

b) Where do you mainly work? City (inc. its suburbs) ☐ Town/Village ☐ Rural ☐

Education

Q.27 Please indicate the level of Education you hold. (Tick the boxes that apply to you)

a) GCSE ☐ c) Bachelor degree ☐ e) PhD ☐

b) A-Levels ☐ d) Master degree ☐

Q.28 What is your area of interest/expertise (academically or professionally)? Please list below:

_____

Familiarity with medium

Q.29 a) How familiar are you with topographic maps? (Tick one box)

Not at all familiar 1   2   3   4   5 Very familiar

☐ ☐ ☐ ☐ ☐

b) With which map data are you most familiar? Please state below.

_____

Q.30 What do you use map data primarily for? (Tick any number of boxes that apply to you)

a) Personal use ☐

b) For research purposes ☐

c) Professional use ☐

d) Don't use any ☐

Q.31 How often do you use map data? (Tick one box)

a) Frequently (more than twice a week) ☐

b) Often (more than once a week) ☐

c) Rarely (less than once a month) ☐

d) Never ☐

Q.32 Any additional comments you would like to share?

_____

_____

_____

Thank you very much for your time!

Appendix B

Survey variables and code book

Table 9 summarises the variables and codes used for processing the data collected from the questionnaire survey. These are used to categorise and code the qualitative results for analysis in the statistical analysis software SPSS version 14.0. Table 9 also lists the applied analysis method for each question/ task. Because of the qualitative nature of this study, the analysis is confined to simple frequency and percentage summaries.

Table 14 Survey variables and codes for analysing the questionnaire survey

Task	Q. No.	Question/ Task Description	Question Type	Data Type	Measure Level	Variable Name	Variable Label	Value Labels (Code)	Missing Values	Analysis method
T A S K 1		Map A interpretation according to use		String	Nominal	• A_Res • A_Retail • A_Office • A_Rail • A_Car • A_OutRec • A_Edu • A_Inst • A_ComAcc • A_Storage • A_Industry • A_Water • A_Rel	• Map A – Residential • Map A – Retailing • Map A – Offices • Map A – Railways • Map A – Car parks • Map A – Outdoor recreation • Map A – Educational buildings • Map A – Institutional buildings • Map A – Institutional buildings & communal accommodation • Map A – Storage & Warehousing • Map A – Industry • Map A – Standing Water • Map A – Religious buildings	0=no response 300=not identified 301=interpreted correctly (same term used) 302=interpreted correctly (different term used) 303= interpreted falsely	0 0 0 0 0 0 0 0 0 0 0 0 0	Frequencies
		Map B interpretation according to purpose		String	Nominal	• B_OutRec • B_Res • B_Edu • B_IndRec • B_Retail	• Map B – Outdoor recreation • Map B – Residential • Map B – Educational buildings • Map B – Indoor recreation • Map B – Retailing	0=no response 300=not identified 301=interpreted correctly (same term used) 302=interpreted correctly	0 0 0 0 0	Frequencies

							• B_Inst	• Map B – Institutional buildings	(different term used)	0	
							• B_Office	• Map B – Offices	303= interpreted falsely	0	
							• B_Rail	• Map B – Railways		0	
							• B_Industry	• Map B – Industry		0	
							• B_Storage	• Map B – Storage & Warehousing		0	
							• B_Rel	• Map B – Religious buildings		0	
T A S K 2	1	How did you approach the map interpretation task?	Open-ended	String	Nominal		• app1 • app2 • app3	• Interpretation approach 1 • Interpretation approach 2 • Interpretation approach 3	0=no response 24=residential areas 78=knowledge from using maps before 79=familiarity of urban layouts 80=size 81=arrangement 82=road network 84=fields 85=symmetry 86=shape 87=pattern 88=space around buildings 89=large objects first 90=similarities 91=relationships between objects 92=tried to find things that were thought of as expected 93=clusters of shapes 94=use of one area to identify the one next to it 95=larger buildings next to houses as shops or schools 96=obvious features 97=detailed examination	1 8 12	Classifying and categorising answers; Percentage summaries

							of area by area 98=familiar areas 99=visioned home town		
2	What captured your attention first?	Open-ended	String	Nominal	• att_A • att_B	• What captured attention in MapA • What captured attention in Map B	0=no response 23=park 24=residential area 36= oval track/sports ground 37=golf course 42=open spaces 44=density of buildings 50=small objects 55=large objects 82=road network	1 1	Classifying and categorising answers; Percentage summaries
3	What do you think is the most dominant object throughout each map? And as what did you interpret it?	Open-ended	String	Nominal	• dom_obj_A • dom_int_A • dom_obj_B • dom_int_B	• Dominating object (Map A) • Dominating object's interpretation (Map A) • Dominating object (Map B) • Dominating object's interpretation (Map B)	0=no response 22=train station 23=park 24=residential area 32=housing estate 33=terraced streets 34=sports stadium 35=running track 37=golf course 42=open spaces 41=building 43=small buildings 45=blocks 54=straight lines 56=oval shape 82=road network 244=recreation	0 0 0 0	Classifying and categorising answers; Percentage summaries
4	Did you use areas you already interpreted to identify what	Closed	String	Nominal	• interp	• Use of already interpreted areas for identifying further ones	0=no response 10=yes 11=no	0	Binary; percentages

		the surrounding areas were?							
5		Do you believe there is a repeating pattern for each land use type?	Closed	String	Nominal	• pattern	• Repeating land use pattern	0=no response 10=yes 11=no	0	Binary; percentages
6		Would it have helped you if the map showed a bigger area?	Closed	String	Nominal	• context	• Context	0=no response 10=yes 11=no	0	Binary; percentages
7		Did the varying scale of the two maps influence your interpretation? If yes, in what way?	Closed/ Open-ended	String	Nominal	• scale	• Scale influence • Scale influence description	0=no response 10=yes 11=no 12=don't know 148=at smaller scale larger areas relate better on the map 149=smaller scale more to interpret 150=things look different 151=clearer the further away 152=context 153=at smaller scale building harder to interpret what they could be used for 154=size of sports facility	0 13	Percentages
8		Was there anything in the map that you weren't sure about?	Open-ended	String	Nominal	• uncertainty	• Uncertainties in map interpretation	0=no response 46=large areas/buildings 47=unfamiliar objects 48=lack of repeating pattern 49=what edges of lines	5	Classifying and categorising answers; Percentage summaries

							represented 155=shops and amenities 156=golf course may be a park 157=uncertain between hospital, housing estate and campus 158=difficult to identify exact structures 159=a lot of areas could be interpreted as a lot of different things 160=at first complex and difficult, but once broken down it was easy		
9	What did you find difficult during your interpretation?	Open-ended/ Closed	String	Nominal	• diff_desc • diff_fuzz • diff_misint • diff_obj_ mean • diff_area_ mean	• Difficulties during map interpretation • Difficulty caused by fuzziness • Difficulty caused by misinterpretation • Difficulty caused by not understanding objects' meaning • Difficulty caused by not understanding areas' meaning	0=no response 10=yes 11=no 47=unfamiliar objects 159=a lot of areas could be interpreted as a lot of different things 161=deciding which area best suits 162=colouring & symbols would be more easier to understand 163=difficult to identify public use buildings 164=identifying more detailed purpose for buildings 165=difference between business and residential 166=interpretation of objects in relation to	2 0 0 0 0	Percentages

						other areas 167=some small areas & shapes between rows of houses difficult to identify 168=difficulties in differentiating between industrial / commercial buildings 169=identifying anything other than open areas and residential areas was a guess		
10	Did you recognise any of the locations shown in the maps? If yes, where do you think the areas are located?	Closed/ Open-ended	String	Nominal	• rec • rec_desc	• Recognition of map's depicted location • Description of map's depicted location	0=no response 10=yes 11=no	0 18	Binary; percentages
11	Which pieces of information do you think are superior to others in the interpretation process?	Attitude	String	Ordinal	• pos • prox • shp • sz • sim_arr • sim_geom • contrast • symm • fact_contxt • simp • orient • conn • struct • clos • comm_reg	• Position • Proximity • Shape • Size • Similarity in arrangement • Similarity in geometry • Contrast • Symmetry • Context • Simplicity of identification • Orientation • Connectedness • Structure • Closure (bound, limit) • Common region	0=no response 1=not important 2 3 4 5=very important	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Frequencies; row proportions

					• cont	• Good continuation of objects		0	
					• likelihood	• Likelihood of correct interpretation		0	
					• org	• Organisation		0	
					• top	• Topology		0	
					• comb	• Combination of objects		0	
12	How important do you rate the following principles for grouping objects together?	Attitude	String	Ordinal	• sim_shp • sim_sz • sim_orien • prox_obj • symm_arr • align_obj • shp_group • rel_part_ whole • dom_obj_ infl	• Similarity in shape • Similarity in size • Similarity in orientation • The proximity between objects • Symmetry in the arrangement • Alignment of objects • The overall shape of the group • The relation among parts and wholes • The influence of one dominating feature within the group	0=no response 1=not important 2 3 4 5=very important	0 0 0 0 0 0 0 0 0	Frequencies; row proportions
13	Did you apply any other grouping principles for you interpretation?	Open-ended	String	Nominal	• group_princ	• Other grouping principle	0=no response 170=distribution of roads 171=some types of building use are more likely to neighbour each other 172=grouping principles really only applied to residential areas	15	Classifying and categorising answers; Percentage summaries
14	How important do you rate the following for your ability to interpret the maps?	Attitude	String	Ordinal	• exist_know • mem • exp • aw • know_frm • know_bel	• Existing knowledge about land use • Memory • Experience • Awareness of our everyday surroundings • Knowledge about the form and sizes of objects constituting a land use • Knowledge of what belongs to a land use category	0=no response 1=not important 2 3 4 5=very important	0 0 0 0 0 0	Frequencies; row proportions

| | 15 | Imagine yourself in the land use you have chosen. What geographical objects make up the land use? | Open-ended | String | Nominal | • LU
• mem_cat1
• mem_cat2
• mem_cat3
• mem_cat4
• mem_cat5 | • Chosen land use concept
• Member category 1
• Member category 2
• Member category 3
• Member category 4
• Member category 5 | 0=no response
14=factory
20=school
21=hospital
22=train station
23=park
24=residential
25=industrial
26=commercial
27=outdoor recreation
28=terraced housing
29=semi-detached housing
30=detached housing
31=public spaces
38=rear garden
39=front garden
40=house
43=small building
210=car park
211=stores/warehouse
212=office
213=open area
214=routes
215=shops
216=main ward building
217=subordinate department building
218=gym
219=school building
220=sport field
221=garden
222=grass
223=pathways
224=trees
225=fence/hedge | 3
3
3
3
8
11 | Frequencies |
| T
A
S
K

3 | | | | | | | | | | |

								227=teaching facilities 228=play ground 229=bus stop 230=restaurant 231=train tracks 232=toilet facilities 233=classroom 234=station building 235=water 236=game facilities		
	16	How do they physically look like in the real world?	Open-ended	String	Nominal	• mem_obj1 • mem_obj2 • mem_obj3 • mem_obj4 • mem_obj5	• Member object 1 • Member object 2 • Member object 3 • Member object 4 • Member object 5	0=no response 19=lake 41=building 42=open space 271=steel lines 272=enclosed space 273=green space 274=multi-storey 275=tarmac space 223=pathways 224=trees 225=fence/hedge	3 3 3 8 11	
	17	What is their primary purpose	Open-ended	String	Nominal	• purp1 • purp2 • purp3 • purp4 • purp5	• Purpose 1 • Purpose 2 • Purpose 3 • Purpose 4 • Purpose 5	0=no response 212=office 215=shops 240=teaching 241=storage 242=sports 243=living accommodation 244=recreation 245=education 246=parking 247=play 248=learning 249=revenue generation	3 3 3 8 11	Semantic relatedness

							250=assist movement 251=aesthetics 252=relaxation 253=buying 254=eating 255=refreshment 256=pick up/drop off 257=walking 258=landscaping 259=accommodating patients 260=treating of patients 261=public seating 276=transport/travel 277=forms railway network		
18	What are they used for?	Open-ended	String	Nominal	• role1 • role2 • role3 • role4 • role5	• Role 1 • Role 2 • Role 3 • Role 4 • Role 5	0=no response 212=office 240=teaching 241=storage 242=sports 243=living accommodation 244=recreation 245=education 246=parking 247=play 248=learning 250=assist movement 251=aesthetics 252=relaxation 253=buying 254=eating 256=pick up/drop off 257=walking 258=landscaping	3 3 3 8 11	Semantic relatedness

							259=accommodating patients 262=sports events 263=treating sick/injured 264=break time 265=school run 266=peeing 267=customer parking 268=entrance to house 269=gardening 270=staff parking 276=transport/travel		
19	How do people make use of them?	Open-ended	String	Nominal	• aff1 • aff2 • aff3 • aff4 • aff5	• Affordance 1 • Affordance 2 • Affordance 3 • Affordance 4 • Affordance 5	0=no response 240=teaching 241=storage 244=recreation 245=education 246=parking 247=play 248=learning 249=revenue generation 252=relaxation 253=buying 254=eating 256=pick up/drop off 257=walking 266=peeing 269=gardening 270=staff parking 276=transport/travel 280=working 281=maintenance 282=driving 283=living 284=curing 285=shopping	4 4 4 9 11	Semantic relatedness

							286=leaving cars 287=meeting 288=shelter 289=exercise 290=sleeping 291=club use 292=enjoyment 293=view 294=shade 295=long stay 296=short visits	
20	What are their geographical properties in terms of geometry, shape, relative size?	Open-ended	String	Nominal	• prop1 • prop2 • prop3 • prop4 • prop5	• Spatial property 1 • Spatial property 2 • Spatial property 3 • Spatial property 4 • Spatial property 5	0=no response 15=linear 16=round 17=rectangular 18=irregular 51=large 52=small 53=square 54=straight lines 56=oval shape 57=spaced out 58=very large 59=very small 60=any shape	3 3 3 8 11
21	Is the member category a kind of or part of your chosen land use? What is the spatial combination of member categories? What is the	Open-ended	String	Nominal	• rel_part_tax1 • rel_part_tax2 • rel_part_tax3 • rel_part_tax4 • rel_part_tax5 • rel_top1 • rel_top2 • rel_top3 • rel_top4 • rel_top5 • rel_dist1	• Taxonomy/Partonomy relation 1 • Taxonomy/Partonomy relation 2 • Taxonomy/Partonomy relation 3 • Taxonomy/Partonomy relation 4 • Taxonomy/Partonomy relation 5 • Topological relation 1 • Topological relation 2 • Topological relation 3 • Topological relation 4 • Topological relation 5 • Distance relation 1	0=no response 61=kind of 62=part of 63=disjoint 64=meet 65=overlap 66=covers 67=covered by 68=contains 69=equal 70=inside	5 5 5 9 12 5 5 5 10 12 4

		spatial distance among member categories?				• rel_dist2 • rel_dist3 • rel_dist4 • rel_dist5	• Distance relation 2 • Distance relation 3 • Distance relation 4 • Distance relation 5	71=zero 72=very close 73=close 74=far	4 4 10 12	
T A S K 4	22	Age group	Closed	String	Nominal	• age	• Age group	0=no response 100=18-25 101=26-30 102=31-35 103=36-40 104=41+	0	Frequencies/ percentage
	23	Gender	Closed	String	Nominal	• gender	• Gender	0=no response 6=male 7=female	0	Frequencies/ percentage
	24	Nationality	Open-ended	String	Nominal	• nat	• Nationality	0=no response 105=British 106=German 107=Brazilian 108=French 109=Portuguese 110=Greek	0	Frequencies/ percentage
	25	Languages spoken (native first)	Open-ended	String	Nominal	• lang1 • lang2 • lang3	• Language spoken 1 • Language spoken 2 • Language spoken 3	0=no response 111=English 112=German 113=Portuguese 114=French 115=Spanish 116=Greek 117=Finnish	0 12 15	Frequencies/ percentage
	26	Where do you live? Where do you work?	Closed	String	Nominal	• live • work	• Place of living • Place of work	0=no response 120=City 121=Town/ Village 122= Rural	0 0	Frequencies/ percentage
	27	Please indicate the level of Education you hold.	Closed	String	Nominal	• gcse • alevel • bsc • msc	• Educational level • Educational level • Educational level • Educational level	0=no response 8=marked 9=unmarked	0 0 0 0	Frequencies/ percentage

					• phd	• Educational level		0	
28	What is your area of interest/ expertise?	Open-ended	String	Nominal	• expert	• Interest/Expertise	0=no response 180=Military mapping 181=communications 182=Armed forces 183=Modern languages/international relations 184=Remote sensing 185=Physics 186=GIS 187=sciences 188=procurement 189=teaching 190=Law	0	Frequencies/ percentage
29	How familiar are you with topographic maps? With which map data are you most familiar?	Attitude / Open-ended	String	Ordinal / Nominal	• familiarity • fam_desc1 • fam_desc2 • fam_desc3	• Familiarity with topographic maps • Familiarity with other map data1 • Familiarity with other map data2 • Familiarity with other map data3	0=no response 1=not at all familiar 2 3 4 5=very familiar 200=Ordnance Survey maps 201=Aerial photography 202=Town plans 203= street maps 204=digital maps 205=terrain models 206=internet based maps (e.g. multimap.com)	0 0 15 17	Frequencies/ percentage
30	What do you use map data primarily for?	Closed	String	Nominal	• use_pers • use_res • use_prof • use_none	• Type of use of map data • Type of use of map data • Type of use of map data • Type of use of map data	0=no response 8=marked 9=unmarked	0 0 0 0	Frequencies/ percentage
31	How often do you use map	Closed	String	Nominal	• frequency	• frequency of map data use	0=no response 140=frequently	0	Frequencies/ percentage

		data?						141=often 142=rarely 143=never		
	32	Any additional comments?	Open-ended	String	Nominal	• comment	• Additional comment to survey	0=no response 400=some questions confusing 401=difficult identifying areas which are not residential/parks 402=confused about the 'purpose' or 'use' of land 403=good survey 404=difficult survey	14	Classifying and categorising answers
				Date		• Date	• Date of completion		0	None
				String		• ID	• Respondent ID		0	None

Appendix C

Description logic preliminaries

Description logics (DL) organise knowledge into classes, gathering the common properties of the data, and the classes themselves into hierarchies. Important characteristics of DL are high expressivity together with decidability and completeness, which guarantee that reasoning algorithms always terminate with the correct answers. They are equipped with several reasoning mechanisms for different types of deduction. They capture the basic facets of data semantics, including the structure of complex entities and ontological dimensions such as time, space, and events.

Description logic family

The smallest propositionally closed DL is $\mathcal{AL}$. It is an attributive language that allows atomic negation ($\neg$), concept intersection ($\sqcap$), universal restriction ($\forall$), and limited existential quantification ($\exists$). In literature, a naming convention is used to describe additional operators and expressivity by extending the base language with the following letters:

$\mathcal{C}$	Complex concept negation.
$\mathcal{S}$	An abbreviation for $\mathcal{AL}$ and $\mathcal{C}$ with transitive properties.
$\mathcal{H}$	Role hierarchy.
$\mathcal{R}$	Limited complex role inclusion axioms; reflexivity and irreflexivity; role disjointness.
$\mathcal{O}$	Nominals (enumerated classes of object value restrictions).
$\mathcal{I}$	Inverse properties.
$\mathcal{N}$	Cardinality restrictions.
$\mathcal{Q}$	Qualified cardinality restrictions.
$\mathcal{F}$	Functional properties.
$\mathcal{E}$	Full existential qualification.
$\mathcal{U}$	Concept union.
$(\mathcal{D})$	Use of data type properties, data values or data types.
$\mathcal{FL}^-$	A sub-language of $\mathcal{AL}$ that prohibits atomic negations.
$\mathcal{FL}_0$	A sub-language of $\mathcal{FL}^-$ that additionally disallows limited existential quantification.

Description logic syntax

The syntax of a language defines the way in which basic elements of the language may be put together to form clauses of that language. Therefore, to reveal the internal structure of a proposition, the sentence must be broken down into smaller parts that can be represented separately. The syntax of description logic consists of a set of unary predicate symbols that are used to denote concept names, a set of binary relations that are used to denote role names, and a recursive definition for defining complex concept terms from concept names and role names using constructors (such as intersection, union, negation, value restrictions, etc.). Generally, concepts denote sets of individuals, and roles denote binary relations between individuals. We differentiate between atomic concepts and complex concept descriptions. Atomic concepts and atomic roles are elementary descriptions, whereas complex descriptions can be built from them inductively with constructors.

The syntax of propositional logic is the most basic form and consists of a countable alphabet Σ of atomic propositions a, b, c, etc. An example of propositional formulas is given below, where | indicates on which side of an axiom a formula can take the following form:

$\phi, \psi \rightarrow a$	Atomic formula
$\mid \top$	Logical truth
$\mid \bot$	Logical falsity
$\mid \neg\phi$	Negation, opposite or complement of sets of individuals
$\mid \phi \wedge \psi$	Conjunction, and-operator, intersection of individuals
$\mid \phi \vee \psi$	Disjunction, or-operator, union of individuals
$\mid \phi \rightarrow \psi$	Implication, conditional, if- or if…then-operator
$\mid \phi \leftrightarrow \psi$	Equivalence, if-and-only-if- or iff-operator

In propositional logic, atomic formulas consist only of concepts. However, often it is difficult to assign a variable to a whole statement. In first-order logic (FOL), the atomic formulas are interpreted as statements about relationships between objects. Since an interpretation $\mathcal{I}$ respectively assigns to every atomic concept and role a unary and binary relation over $\Delta^{\mathcal{I}}$, we can view atomic concepts and roles as unary and binary predicates. Then, any concept C can be translated effectively into a predicate logic

formula $\phi_C(x)$ with one free variable x such that for every interpretation $\mathcal{I}$ the set of elements $\Delta^{\mathcal{I}}$ satisfying $\phi_C(x)$ is exactly $C^{\mathcal{I}}$. An atomic concept A is translated into the formula $A(x)$ (Baader *et al.*, 2003). For example, $C \sqcap D$ can be regarded as the first order logic sentence $C(x) \wedge D(x)$, where the variable x ranges over all individuals in the interpretation domain, and $C(x)$ is true for those individuals that belong to the concept C. Therefore, in FOL a well-formed formula is defined as $P(t_1, \ldots, t_n)$ where P is an n-ary predicate of a language $\mathcal{L}$ consisting of a set U, and each of $t_1, \ldots, t_n$ is either a variable or an element of U. First-order logic is a two-valued logic with just two quantifiers and the basic Boolean operators introduced above, thus extending the terms of propositional logics as follows:

Terms:	t $\rightarrow$ x	Variable
	| a	Constant
	| f (t$_1$, ..., t$_n$)	Function application
Formulas:	$\psi, \phi \rightarrow$ P(t$_1$, ..., t$_n$)	Atomic formulas
	...	...
	| $\forall$x.ϕ	Universal quantification
	| $\exists$x.ϕ	Existential quantification

Description logic semantics

In DL, semantics is defined by interpreting concepts as sets of individuals and roles as sets of pairs of individuals (Baader *et al.*, 2003). The model-theoretic semantics of a logic is given in the standard form using a Tarskian interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ consisting of a non-empty set $\Delta^{\mathcal{I}}$ (the domain of the interpretation), and an interpretation function $\cdot^{\mathcal{I}}$ that maps

- Every concept to a subset of $\Delta^{\mathcal{I}}$: Concept name $A \rightarrow$ subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ ($A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$).
- Every role to a subset of $\Delta^{\mathcal{I}}$ x $\Delta^{\mathcal{I}}$: Role name $R \rightarrow$ binary relation $R^{\mathcal{I}}$ over $\Delta^{\mathcal{I}}$ ($R^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$).
- Every individual to an element of $\Delta^{\mathcal{I}}$: Individual name $i \rightarrow i^{\mathcal{I}}$ element of $\Delta^{\mathcal{I}}$ ($i^{\mathcal{I}} \in \Delta^{\mathcal{I}}$).

This means for a given set as the domain, an interpretation of $\mathcal{AL}$ concepts is defined as an atomic concept when it is interpreted as a set of individuals that is a subset of the domain. It is interpreted as an atomic role when it is interpreted as a set of pairs of individuals from the domain, i.e., a binary relation over the domain. For example, in a

given interpretation $\mathcal{I}$, whose domain contains the element a, we have that $a \in (C \sqcap D)^{\mathcal{I}}$, then from the semantics we know that such element should be in the intersection of $C^{\mathcal{I}}$ and $D^{\mathcal{I}}$, that is, it should be in both $C^{\mathcal{I}}$ and $D^{\mathcal{I}}$. In an interpretation, different individuals are assumed to denote different elements, that is, for every pair of individuals a, b, and for every interpretation $\mathcal{I}$, if $a \neq b$ then $a^{\mathcal{I}} \neq b^{\mathcal{I}}$. This is called the unique name assumption and is usually assumed in database applications (Donini *et al.*, 1996).

Therefore, if a truth value assignment (or interpretation) of the atoms in Σ is a function $\mathcal{I}: \Sigma \rightarrow \{T, F\}$, then a formula ϕ is satisfied by an interpretation $\mathcal{I}$ ($\mathcal{I} \vDash \phi$) or is true under $\mathcal{I}$ when:

$\mathcal{I} \vDash \top$

$\mathcal{I} \nvDash \bot$

$\mathcal{I} \vDash a$ iff $a^{\mathcal{I}} = T$

$\mathcal{I} \vDash \neg\phi$ iff $\mathcal{I} \nvDash \phi$

$\mathcal{I} \vDash \phi \wedge \psi$ iff $\mathcal{I} \vDash \phi$ and $\mathcal{I} \vDash \psi$

$\mathcal{I} \vDash \phi \vee \psi$ iff $\mathcal{I} \vDash \phi$ or $\mathcal{I} \vDash \psi$

$\mathcal{I} \vDash \phi \rightarrow \psi$ iff $\mathcal{I} \vDash \phi$, then $\mathcal{I} \vDash \psi$

$\mathcal{I} \vDash \phi \leftrightarrow \psi$ iff $\mathcal{I} \vDash \phi$, if and only if $\mathcal{I} \vDash \psi$

Where $\vDash$ means implies, $\nvDash$ not implies, and iff means if and only if. For example, let C and D be concept descriptions and A be an atomic concept of the language $\mathcal{AL}$, then the top concept is interpreted as the whole domain (i.e., $\top^{\mathcal{I}} = \Delta^{\mathcal{I}}$). The bottom concept is interpreted as the empty set (i.e., $\bot^{\mathcal{I}} = \emptyset$). The interpretation of $\neg A$ is the set of all individuals in the domain that do not belong to the interpretation of A (i.e., $\Delta^{\mathcal{I}} \setminus A^{\mathcal{I}}$). Intersection of two concepts is interpreted as the set-intersection of all individuals in the domain that belong to both interpretation of C and the interpretation of D (i.e., $C^{\mathcal{I}} \cap D^{\mathcal{I}}$). The union of concepts means that individuals in the domain are instances of either C or D, and is interpreted as $(C \sqcup D)^{\mathcal{I}} = C^{\mathcal{I}} \cup D^{\mathcal{I}}$. Two concepts are equivalent $C \equiv D$ if and only if $C^{\mathcal{I}} = D^{\mathcal{I}}$ for all interpretations $\mathcal{I}$ (Baader *et al.*, 2003), that is, the individuals in the domain of C are equivalent to the individuals of D.

An interpretation is uniquely determined by the values that it gives to primitive concepts and primitive roles. Hence, a complex sentence has a meaning derived from the meaning from its parts. A sentence is valid or necessarily true if and only if it is true under all possible interpretations in all possible worlds. Such sentences are referred to as tautologies. A sentence is satisfiable if and only if there is some interpretation in some world for which it is true (Russell and Norvig, 1995). An interpretation is satisfiable and valid under the following conditions: An interpretation $\mathcal{I}$ is a model of ϕ ($\mathcal{I} \vDash \phi$). A formula ϕ is

- Satisfiable, if there is some $\mathcal{I}$ that satisfies ϕ.

- Unsatisfiable, if ϕ is not satisfiable.

- Falsifiable, if there is some $\mathcal{I}$ that does not satisfy ϕ.

- Valid, if every $\mathcal{I}$ is a model of ϕ (tautology).

- Two formulas are logically equivalent ($\phi \equiv \psi$), if for all $\mathcal{I}$: $\mathcal{I} \vDash \phi$ iff $\mathcal{I} \vDash \psi$.

As in propositional logic, a complex FOL formula may be true or false with respect to a given interpretation. The interpretation specifies referents for constant symbols as objects, for predicate symbols as relations, and for function symbols as functional relations. For example, an atomic sentence $P(t_1, \ldots, t_n)$ is true in a given interpretation iff the objects referred to by $t_1, \ldots, t_n$ are in the relation referred to by the predicate P. Therefore, the interpretation $\mathcal{I} = \langle \Delta, \cdot^{\mathcal{I}} \rangle$ is an arbitrary non-empty set Δ and $\cdot^{\mathcal{I}}$ is a function that maps n-ary function symbols over Δ ($f^{\mathcal{I}} \in [\Delta^n \to \Delta]$), individual constants to elements of Δ ($a^{\mathcal{I}} \in \Delta$), and n-ary predicate symbols to relation over Δ ($P^{\mathcal{I}} \subseteq \Delta^n$). A formula ϕ is satisfied by (is true in) an interpretation $\mathcal{I}$ under a variable assignment α ($\mathcal{I}, \alpha \vDash \phi$):

$\mathcal{I}, \alpha \vDash P(t_1, \ldots, t_n)$ iff $\langle t_1^{\mathcal{I},\alpha}, \ldots, t_n^{\mathcal{I},\alpha} \rangle \in P^{\mathcal{I}}$

$\mathcal{I}, \alpha \vDash \neg\phi$ iff $\mathcal{I}, \alpha \nvDash \phi$

$\mathcal{I}, \alpha \vDash \phi \wedge \psi$ iff $\mathcal{I}, \alpha \vDash \phi$ and $\mathcal{I}, \alpha \vDash \psi$

$\mathcal{I}, \alpha \vDash \phi \vee \psi$ iff $\mathcal{I}, \alpha \vDash \phi$ or $\mathcal{I}, \alpha \vDash \psi$

$\mathcal{I}, \alpha \vDash \forall x.\phi$ iff for all $d \in \Delta$: $\mathcal{I}, \alpha[x/d] \vDash \phi$

$\mathcal{I}, \alpha \vDash \exists x.\phi$ iff there exists $d \in \Delta$: $\mathcal{I}, \alpha \ [x/d] \vDash \phi$

Similar as in propositional logic, a formula ϕ can be satisfiable, unsatisfiable, falsifiable or valid, except that the definition is in terms of the pair $(\mathcal{I}, \alpha)$. For example, a formula ϕ is satisfiable if there is some $(\mathcal{I}, \alpha)$ that satisfies ϕ, and it is valid if every $(\mathcal{I}, \alpha)$ is a model of ϕ. Overall, an interpretation function $\cdot^{\mathcal{I}}$ is an extension function if and only if it satisfies the semantic definitions of the language.

Knowledge bases

A knowledge base $\mathcal{KB}$ is a pair $\langle \mathcal{T}, \mathcal{A} \rangle$ where $\mathcal{T}$ is a set of "terminological" axioms and $\mathcal{A}$ is a set of "assertional" axioms: $\Sigma = \langle \text{TBox}, \text{ABox} \rangle$. Terminological axioms in the TBox are restricted to so-called definitions, where a definition is an assertion stating that the extension of a concept denoted by a name is equal to the extension of another complex concept (Calvanese *et al.*, 2001). A concept name A directly uses a concept name B in a TBox Σ iff the definition of A mentions B. A concept A uses a concept name B_n iff there is a chain of concept names $\langle A, B_1, \ldots, B_n \rangle$ such that B_i directly uses B_{i+1}. A TBox is acyclic iff no concept name uses itself. For example, if we build a graph whose nodes are atomic concepts and whose arcs connect pairs of concepts such that one appears in the definition of the other, then the graph is acyclic.

The ABox is a set $\mathcal{A}$ of assertions that is realised by permitting concepts and roles to be used in assertions on individuals. These assertional axioms state the ground facts of the $\mathcal{KB}$. Given a concept language $\mathcal{L}$, an ABox-statement in $\mathcal{L}$ has either one of the two forms:

- $C(a)$ Concept membership assertion
- $R(a, b)$ Role membership assertion

where a, b are individual names, C is a concept name and R is a role name.

Different semantics have been proposed for the TBox depending on the fact whether cyclic statements are allowed or not (Donini *et al.*, 1996). In descriptive semantics, an interpretation $\mathcal{I}$ satisfies (models) a TBox axiom A ($\mathcal{I} \vDash A$):

$\mathcal{I} \vDash C \sqsubseteq D$ iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$

$\mathcal{I} \vDash R \sqsubseteq S$ iff $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$

$\mathcal{I} \vDash R^+ \sqsubseteq R$ iff $(R^{\mathcal{I}})^+ \subseteq R^{\mathcal{I}}$

$\mathcal{I} \vDash C \equiv D$ iff $C^{\mathcal{I}} = D^{\mathcal{I}}$

$\mathcal{I} \vDash R \equiv S$ iff $R^{\mathcal{I}} = S^{\mathcal{I}}$

$\mathcal{I}$ satisfies a Tbox $\mathcal{T}$ ($\mathcal{I} \vDash \mathcal{T}$) iff $\mathcal{I}$ satisfies every axiom A in $\mathcal{T}$, that is, an interpretation $\mathcal{I}$ is a model for a TBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all the statements in $\mathcal{T}$. An interpretation $\mathcal{I}$ satisfies a role hierarchy $\mathcal{H}$ if it satisfies all role inclusions ($\mathcal{I} \vDash R \sqsubseteq S$ iff $R^{\mathcal{I}} \subseteq S^{\mathcal{I}}$) in $\mathcal{H}$. The above definitions are constraints stating a restriction on the valid models of the knowledge base, and in particular on the possible interpretations of a given concept or role, where it is no better specified. Although descriptive semantics seems most appropriate, it is not satisfactory in all cases, such as in cyclic statements (Donini *et al.*, 1996).

The semantics of an ABox is defined as follows: Given a set $\mathcal{A}$ of assertions and if $\mathcal{I}=(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is an interpretation, then C(a) is satisfied by $\mathcal{I}$ if $a^{\mathcal{I}} \in C^{\mathcal{I}}$, and R($a$, b) is satisfied by $\mathcal{I}$ if $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in R^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is said to be a model of the ABox $\mathcal{A}$ if every assertion of $\mathcal{A}$ is satisfied by $\mathcal{I}$, that is, $\mathcal{I}$ satisfies an Abox $\mathcal{A}$ ($\mathcal{I} \vDash \mathcal{A}$) iff $\mathcal{I}$ satisfies every axiom A in $\mathcal{A}$ (Donini *et al.*, 1996).

An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is said to be a model of a knowledge base Σ if every axiom of Σ is satisfied by $\mathcal{I}$, that is, $\mathcal{I}$ satisfies a $\mathcal{KB}$ $\mathcal{K}$ ($\mathcal{I} \vDash \mathcal{K}$) iff $\mathcal{I}$ satisfies both $\mathcal{T}$ and $\mathcal{A}$. A knowledge base $\Sigma= \langle$TBox, ABox$\rangle$ is said to be satisfiable if it admits to a model. In particular, satisfiability of concept terms can be reduced to ABox consistency as follows: A concept term C is satisfiable iff the ABox $\{C(a)\}$ is consistent (Haarslev and Möller, 2000). Therefore, an ABox is consistent with respect to a TBox $\mathcal{T}$ iff it has a model $\mathcal{I}$ that is also a model of $\mathcal{T}$.

Reasoning

A model m of a sentence α is true if α is true in m. If $M(\alpha)$ is the set of all models of α, then knowledge base $\mathcal{KB} \vDash \alpha$ if and only if $M(\mathcal{KB}) \subseteq M(\alpha)$. The model refers to the interpretation of logical statements. The logical implication $\mathcal{KB} \vDash \alpha$ means that $\mathcal{KB}$ entails sentence α if and only if α is true in all worlds where $\mathcal{KB}$ is true (Russell and

Norvig, 1995). A formula ϕ can be implied by sets of formulas Θ if ϕ is true in all models of Θ, that is, $\Theta \vDash \phi$ iff $\mathcal{I} \vDash \phi$ for all models $\mathcal{I}$ of Θ. Let $\mathcal{KB}$ ($\Sigma = \langle$TBox, ABox$\rangle$) be a knowledge base, $\mathcal{A}$ be an ABox, $\mathcal{T}$ be a TBox, C and D concept descriptions, and a an individual name, then reasoning services can be described as follows (Calvanese *et al.*, 2001; Baader and Küsters, 2006):

- Subsumption ($\Sigma \vDash C \sqsubseteq D$) is the problem of checking whether C is subsumed by D with respect to Σ, that is, whether $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ is in every model $\mathcal{I}$ of Σ.

- Satisfiability ($\Sigma \nvDash$) is the problem of checking whether Σ has a model. Concept satisfiability ($\Sigma \nvDash C \equiv \bot$) is a special case of subsumption, with the subsumer being the empty concept, meaning that a concept is not satisfiable. Concept satisfiability is therefore the problem of checking whether C is satisfiable with respect to Σ, that is, whether there exists a model $\mathcal{I}$ of Σ such that $C^{\mathcal{I}} \neq \emptyset$. Knowledge base satisfiability is the problem of deciding whether a knowledge base $\mathcal{KB}$ is satisfiable, that is, whether $\mathcal{KB}$ admits a model $\mathcal{I}$.

- Concept consistency is the problem of deciding whether a concept C is consistent in a knowledge base $\mathcal{KB}$, that is, whether $\mathcal{KB}$ admits a model $\mathcal{I}$ such that $C^{\mathcal{I}} \neq \emptyset$. The ABox consistency problem is to decide whether a given ABox $\mathcal{A}$ is consistent with respect to a TBox $\mathcal{T}$.

- Logical implication is the problem of deciding whether a knowledge base $\mathcal{KB}$ implies an inclusion assertion $C_1 \sqsubseteq C_2$ (written $\mathcal{KB} \vDash C_1 \sqsubseteq C_2$), that is, whether $C_1^{\mathcal{I}} \subseteq C_2^{\mathcal{I}}$ for each model $\mathcal{I}$ of $\mathcal{KB}$.

- Equivalence of concepts within a terminology $\mathcal{T}$ is deciding whether two concepts are logically equivalent ($C \equiv_{\mathcal{T}} D$), that is, if in all models $\mathcal{I}$ of $\mathcal{T}$ we have $C^{\mathcal{I}} = D^{\mathcal{I}}$.

- Instance checking ($\Sigma \vDash C(a)$) is the problem of checking whether the assertion $C(a)$ is satisfied in every model of Σ.

- Retrieval ($\{a \mid \Sigma \vDash C(a)\}$) is the problem of checking whether an individual occurring in the assertions is an instance of some concept description C.

- Realisation ($\{C \mid \Sigma \vDash C(a)\}$) is the problem of checking the most specific concept C in the TBox that a is an instance of.

Inference rules rely implicitly on a general property of certain logics (e.g. propositional and first-order logic) called monotonicity. A logic is monotonic if when we add some new sentence to the knowledge base, all the sentences entailed by the original $\mathcal{KB}$ are still entailed by the new larger knowledge base. Formally, this is expressed as if $\mathcal{KB}_1 \vDash \alpha$ then $(\mathcal{KB}_1 \cup \mathcal{KB}_2) \vDash \alpha$. Were it not for monotonicity, we could not have any local inference rules because the rest of the $\mathcal{KB}$ might affect the soundness of the inference. This would potentially cripple any inference procedure (Russell and Norvig, 1995). Therefore, given that sentence α can be derived from the set of sentences $\mathcal{KB}$ by procedure i ($\mathcal{KB} \vdash i \; \alpha$), then:

- Procedure i is sound whenever procedure i proves that a sentence α can be derived from a set of sentences $\mathcal{KB}$ ($\mathcal{KB} \vdash i \; \alpha$), then it is also true that $\mathcal{KB}$ entails α ($\mathcal{KB} \vDash \alpha$). This means that no wrong inferences are drawn, although a sound procedure may fail to find the solution in some cases, where there is actually one.

- Procedure i is complete whenever a set of sentences $\mathcal{KB}$ entails a sentence α ($\mathcal{KB} \vDash \alpha$), then procedure i proves that α can be derived from $\mathcal{KB}$ ($\mathcal{KB} \vdash i \; \alpha$).

  This means all the correct inferences are drawn, but a complete procedure may claim to have found a solution in some cases, when there is actually no solution.

Concrete domains

The idea of adding data type properties and data values is based on capturing the concrete semantics of objects not with description logic axioms, but to represent them separately (Esposito *et al.*, 2007). A concrete domain $\mathcal{D}$ is a tuple ($\Delta^{\mathcal{D}}$, Φ) of a non-empty set $\Delta^{\mathcal{D}}$ and a set of predicates Φ. Each predicate name $P_{\mathcal{D}}$ from $\Phi_{\mathcal{D}}$ is associated with an arity n and an n-ary predicate $P_{\mathcal{D}} \subseteq \Delta_{\mathcal{D}}{}^{n}$. Attributes are introduced as partial functions that map individuals of the abstract domain $\Delta^{\mathcal{I}}$ to elements of $\Delta^{\mathcal{D}}$ of the concrete domain $\mathcal{D}$. For attributes a, the interpretation is extended as $a^{\mathcal{I}} \colon \Delta^{\mathcal{I}} \to \Delta^{\mathcal{D}}$. Concrete domains are restricted to so-called admissible concrete domains to keep the inference problems of this extension decidable.

Ontologies

The interpretation of an ontology is defined as the collection of all the legal world descriptions that conform to the constraints imposed by the ontology. To formally define the interpretation, an ontology is mapped into a set of first-order logic formulas. The legal world descriptions (i.e. the interpretation) of an ontology are all the models of the translated FOL theory. Therefore, given an interpretation $\mathcal{I} = \langle D, \cdot^{\mathcal{I}} \rangle$ where D is an arbitrary non-empty set such that $D = \Omega \cup \mathcal{B}$, where $\mathcal{B} = \bigcup_{i=1}^{m} \mathcal{B}_{Di}$. $\mathcal{B}_{Di}$ is the set of values associated with each basic domain (i.e. integer, string, etc.) and $\mathcal{B}_{Di} \cap \mathcal{B}_{Dj} = \emptyset$, $\forall i,j.i \neq j$, and Ω is the abstract entity domain such that $\mathcal{B} \cap \Omega = \emptyset$. Then the interpretation function maps:

- Basic domain predicates to elements of the relative basic domain $D_i^{\mathcal{I}} = \mathcal{B}_{Di}$, e.g. string$^{\mathcal{I}} = \mathcal{B}_{string}$;
- Entity-set predicates to elements of the entity domain $E_i^{\mathcal{I}} \subseteq \Omega$;
- Attribute predicates to binary relations such that $A_i^{\mathcal{I}} \subseteq \Omega \times \mathcal{B}$;
- Relationship-set predicates to n-ary relations over the entity domain $R_i^{\mathcal{I}} \subseteq \Omega \times \Omega \dots \times \Omega = \Omega^n$;

where the alphabet of the FOL language will have the predicate symbols $E_1, E_2, \dots, E_n$ for each entity set, $D_1, D_2, \dots, D_m$ for each basic domain, $A_1, A_2, \dots, A_k$ for each attribute, and $R_1, R_2, \dots, R_p$ for each relationship-set.

When data types are supported, the domain is divided into two disjoint sets of the 'object domain' $\Delta^{\mathcal{I}}_{\mathcal{O}}$ and the 'data type' domain $\Delta^{\mathcal{I}}_{\mathcal{D}}$ such that $\Delta^{\mathcal{I}} = \Delta^{\mathcal{I}}_{\mathcal{O}} \cup \Delta^{\mathcal{I}}_{\mathcal{D}}$. The interpretation then maps individuals into elements of the object domain, classes into subsets of the object domain, data types into subsets of the data type domain and data values into elements of the data type domain. Object properties and data type properties are mapped into subsets of $\Delta^{\mathcal{I}}_{\mathcal{O}} \times \Delta^{\mathcal{I}}_{\mathcal{D}}$ respectively. Thus, individuals and data values correspond to FOL constants, classes and data types correspond to unary predicates, properties correspond to binary predicates, and sub-class/property relationships correspond to implication (Horrocks *et al.*, 2003).

Appendix D

This appendix lists all the code used (SQL, Python, OWL) for implementing the proposed conceptual model, as described in chapter 7. This includes the preparation of the database in terms of extracting a sample area and calculating missing relations, the transformation from database instances to OWL individuals, and example OWL code from the ontology and its asserted individuals.

SQL code

Extracting a sample dataset

For the purpose of the proof of concept, we need to extract a small dataset that will be used for the reasoning, that is, to demonstrate the use of description logics to infer higher order information. Using the SQL buffer operator SDO_WITHIN_DISTANCE, data can be extracted within a specified distance of 1000 metres:

```
Create table Glasgow_Sample as
Select b.feature_id, b.toid, b.baseform, b.basefunc, b.location
From maia.gb04_s_ft_topo_area a, maia.gb04_s_ft_topo_area b
Where a.feature_id='{}' and sdo_within_distance(b.location, a.location, 'distance=1000 unit= meter')
= 'TRUE';
```

From the reduced dataset, we then extract only the building features. This is done for two reasons: To calculate the spatial relation between buildings, and to reduce the number of features that the description logic system must handle. The derived dataset is called RESIDENTIAL_BUILDINGS:

```
Create table RESIDENTIAL_BUILDINGS as
Select b.TOID, b.FEATURE_ID, a.FEATURECODE, b.baseform, a.BASEFORM_DESC, a.BFORM_FULL_STRING_FORM,
b.BASEFUNC, a.BASEFUNC_DESC, a.BFUNC_FULL_STRING_FORM, a.OSMMTOPO_DESCRIPTIVE_GROUPS,
a.OSMMTOPO_DESCRIPTIVE_TERMS, a.OSMMTOPO_MAKE, a.OSMMTOPO_THEMES, a.CALCULATEDAREAVALUE,
a.PHYSICALLEVEL, a.PHYSICALPRESENCE, b.LOCATION
From residential a, sample_2_2D b
Where b.TOID = a.TOID (+) and a.OSMMTOPO_DESCRIPTIVE_GROUPS='Building';
```

To be able to visualise the newly created tables, it is important to keep the spatial geometry column. We then have to insert a record in the spatial metadata for the tables based on their geometry column LOCATION:

```
Insert into user_sdo_geom_metadata
values (
'Residential_Buildings',
'Location',
 MDSYS.SDO_DIM_ARRAY (
  MDSYS.SDO_DIM_ELEMENT('X',0,700000,0.001),
  MDSYS.SDO_DIM_ELEMENT('Y',0,1300000,0.001)
  ),
81989);
```

This is followed by creating a spatial index:

```
CREATE INDEX RESIDENTIAL_BUILDINGS_GI ON RESIDENTIAL_BUILDINGS(LOCATION)
    INDEXTYPE IS MDSYS.SPATIAL_INDEX;
```

Calculating missing spatial relations

With a table that consists only of building features, we can now calculate the 'touch' relations between all buildings using the SDO_TOUCH operator:

```
Create table RES_BUILDINGS_TOUCH as
Select a.TOID as TOID_BUILDING1, b.TOID as TOID_BUILDING2
From maiagen.RESIDENTIAL_BUILDINGS a, maiagen.RESIDENTIAL_BUILDINGS b
Where sdo_touch(a.LOCATION, b.LOCATION)='TRUE';
```

By joining the original table RESIDENTIAL_BUILDINGS with the derived RES_BUILDINGS_TOUCH table, we create one table with all the information necessary for the knowledge base:

```
Create table HOUSES_TOUCH as
Select a.*, b.TOID_BUILDING1, b.TOID_BUILDING2
From maiagen.RESIDENTIAL_BUILDINGS a, maiagen.RES_BUILDINGS_TOUCH b
Where a.TOID = b.TOID_BUILDING1 (+);
```

The symbol (+) stands for an OUTER JOIN to include building features that do not touch any other buildings. The table HOUSES_TOUCH carries only relevant columns required for the reasoning. Because of the complexity of the asserted relations between individuals in the knowledge base, this table is reduced further to building features that are of size (CALCULATEDAREAVALUE) greater than $35m^2$. Finally, the table is exported as a comma separated values (CSV) file for further processing.

Python code

Python script is used to read the CSV file and to input its contents into RDF syntax. First, we assert one individual in OWL the way we want all individuals to be asserted. The generated OWL code form this individual can be then used for the python script, where the code is populated with information from the CSV file. This creates a text file of the specified code for each row, that is, every individual from the CSV file.

Example for an individual that touches no other building

Below is the script for individuals that do not touch any other buildings.

```
# import the csv library

import csv

# create a reader object to parse the csv file. Note the double "\\" in the directory
# path. This is important otherwise python won't find your file.

reader =
csv.reader(file("F:\\PhD\\GeoBase04\\ResidentialAnalysis\\Glasgow_Sample_2_2D\\RES_BUILD_HOUSES_JOIN_
TOUCH_0.csv"))

#Step through each row in the csv file

for row in reader:

#The RDF string contains the RDF. The data in the first column of each row can be accessed using
#row[0]. Data in the nth column is accessed using row[n-1]. Single quotes are used to hold the string
#values. The control character \n returns a newline.

    RDF = ('    <!-- http://www.semanticweb.org/ontologies/2008/4/GlasgowSample.owl#' + row[0] + ' --
>\n'

    + '<Building rdf:about="#' + row[0] + '">\n'
        + '<rdf:type>\n'
            + '<owl:Class>\n'
                + '<owl:complementOf>\n'
                    + '<owl:Restriction>\n'
                        + '<owl:onProperty rdf:resource="#touches"/>\n'
                        + '<owl:someValuesFrom rdf:resource="#Building"/>\n'
                    + '</owl:Restriction>\n'
                + '</owl:complementOf>\n'
            + '</owl:Class>\n'
        + '</rdf:type>\n'
        + '<hasArea rdf:datatype="&xsd;float">' + row[2] + '</hasArea>\n'
    + '</Building>\n')

# The print statement below print the value of RDF to the console window. You can then copy/paste
this to a text file or whatever
# Note for some reason you need to use ctrl c/ctrl v for copy and paste in the console window as
right-clicking with the mouse
# won't work.

    print RDF
```

Example for individuals that touch other buildings

For individuals that touch other buildings, the code changes slightly because we have to assert which building the individual touches:

```
    RDF = ('    <!-- http://www.semanticweb.org/ontologies/2008/4/GlasgowSample.owl#' + row[0] + ' --
>\n'

    + '<Building rdf:about="#' + row[0] + '">\n'
        + '<rdf:type>\n'
            + '<owl:Restriction>\n'
                + '<owl:onProperty rdf:resource="#touches"/>\n'
                + '<owl11:onClass rdf:resource="#Building"/>\n'
                + '<owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>\n'
            + '</owl:Restriction>\n'
        + '</rdf:type>\n'
        + '<hasArea rdf:datatype="&xsd;float">' + row[2] + '</hasArea>\n'
        + '<touches rdf:resource="#' + row[4] + '"/>\n'
    + '</Building>\n')
```

For individuals that touch more than 1 building, we simply add rows:

```
        + '<touches rdf:resource="#' + row[5] + '"/>\n'
        + '<touches rdf:resource="#' + row[6] + '"/>\n'
```

OWL code from Protégé

The generated syntax from the Python script can then be loaded into Protégé, where we will now find all asserted individuals. We define the high-level concepts of the TBox as described in chapter 7. Below some of the OWL code is given.

HousesOntology_Test3TouchRelation_allclassified.owl:

The sample dataset contains thousands of individuals, which results in very large OWL files. For illustrative purpose, only the building ontology for figure 38 with only a few individuals is given here:

```
<?xml version="1.0"?>


<!DOCTYPE rdf:RDF [
    <!ENTITY owl "http://www.w3.org/2002/07/owl#" >
    <!ENTITY owl11 "http://www.w3.org/2006/12/owl11#" >
    <!ENTITY xsd "http://www.w3.org/2001/XMLSchema#" >
    <!ENTITY owl11xml "http://www.w3.org/2006/12/owl11-xml#" >
    <!ENTITY rdfs "http://www.w3.org/2000/01/rdf-schema#" >
    <!ENTITY rdf "http://www.w3.org/1999/02/22-rdf-syntax-ns#" >
    <!ENTITY HousesOntology "http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#" >
    <!ENTITY HousesOntology-Test3TouchRelation-terraces
"http://www.semanticweb.org/ontologies/2008/4/HousesOntology-Test3TouchRelation-terraces.owl#" >
]>
```

```xml
<rdf:RDF xmlns="http://www.semanticweb.org/ontologies/2008/4/HousesOntology-Test3TouchRelation-
allclassified.owl#"
    xml:base="http://www.semanticweb.org/ontologies/2008/4/HousesOntology-Test3TouchRelation-
allclassified.owl"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:HousesOntology-Test3TouchRelation-
terraces="http://www.semanticweb.org/ontologies/2008/4/HousesOntology-Test3TouchRelation-
terraces.owl#"
    xmlns:HousesOntology="http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#"
    xmlns:owl11="http://www.w3.org/2006/12/owl11#"
    xmlns:owl11xml="http://www.w3.org/2006/12/owl11-xml#"
    xmlns:owl="http://www.w3.org/2002/07/owl#"
    xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
    xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
 <owl:Ontology rdf:about=""/>



 <!--
 ///////////////////////////////////////////////////////////////////////////////
 //
 // Object Properties
 //
 ///////////////////////////////////////////////////////////////////////////////
  -->




 <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#touches -->

 <owl:ObjectProperty rdf:about="&HousesOntology;touches">
     <rdf:type rdf:resource="&owl;SymmetricProperty"/>
     <owl:inverseOf rdf:resource="&HousesOntology;touches"/>
 </owl:ObjectProperty>




 <!--
 ///////////////////////////////////////////////////////////////////////////////
 //
 // Data properties
 //
 ///////////////////////////////////////////////////////////////////////////////
  -->




 <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#hasArea -->

 <owl:DatatypeProperty rdf:about="&HousesOntology;hasArea">
     <rdf:type rdf:resource="&owl;FunctionalProperty"/>
 </owl:DatatypeProperty>




 <!--
```

terraces="http://www.semanticweb.org/ontologies/2008/4/HousesOntology-Test3TouchRelation-

```
//////////////////////////////////////////////////////////////////////////////////
//
// Classes
//
//////////////////////////////////////////////////////////////////////////////////
 -->



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#Building -->

<owl:Class rdf:about="&HousesOntology;Building">
    <rdfs:subClassOf rdf:resource="&owl;Thing"/>
</owl:Class>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#DetachedHouse -->

<owl:Class rdf:about="&HousesOntology;DetachedHouse">
    <owl:equivalentClass>
        <owl:Class>
            <owl:intersectionOf rdf:parseType="Collection">
                <rdf:Description rdf:about="&HousesOntology;House"/>
                <owl:Class>
                    <owl:complementOf>
                        <owl:Restriction>
                            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                            <owl:someValuesFrom rdf:resource="&HousesOntology;House"/>
                        </owl:Restriction>
                    </owl:complementOf>
                </owl:Class>
            </owl:intersectionOf>
        </owl:Class>
    </owl:equivalentClass>
    <rdfs:subClassOf rdf:resource="&HousesOntology;House"/>
</owl:Class>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#EndTerracedHouse -->

<owl:Class rdf:about="&HousesOntology;EndTerracedHouse">
    <owl:equivalentClass>
        <owl:Class>
            <owl:intersectionOf rdf:parseType="Collection">
                <owl:Class>
                    <owl:complementOf rdf:resource="&HousesOntology;MidTerracedHouse"/>
                </owl:Class>
                <rdf:Description rdf:about="&HousesOntology;House"/>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl:someValuesFrom rdf:resource="&HousesOntology;MidTerracedHouse"/>
                </owl:Restriction>
            </owl:intersectionOf>
        </owl:Class>
    </owl:equivalentClass>
    <rdfs:subClassOf rdf:resource="&HousesOntology;House"/>
```

```
        </owl:Class>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#House -->

    <owl:Class rdf:about="&HousesOntology;House">
        <owl:equivalentClass>
            <owl:Class>
                <owl:intersectionOf rdf:parseType="Collection">
                    <rdf:Description rdf:about="&HousesOntology;Building"/>
                    <owl:Class>
                        <owl:intersectionOf rdf:parseType="Collection">
                            <owl:Restriction>
                                <owl:onProperty rdf:resource="&HousesOntology;hasArea"/>
                                <owl:someValuesFrom>
                                    <rdf:Description>
                                        <rdf:type rdf:resource="&owl;DataRange"/>
                                        <owl11:minInclusive
rdf:datatype="&xsd;int">35</owl11:minInclusive>
                                        <owl11:onDataRange rdf:resource="&xsd;float"/>
                                    </rdf:Description>
                                </owl:someValuesFrom>
                            </owl:Restriction>
                            <owl:Restriction>
                                <owl:onProperty rdf:resource="&HousesOntology;hasArea"/>
                                <owl:someValuesFrom>
                                    <rdf:Description>
                                        <rdf:type rdf:resource="&owl;DataRange"/>
                                        <owl11:maxInclusive
rdf:datatype="&xsd;int">160</owl11:maxInclusive>
                                        <owl11:onDataRange rdf:resource="&xsd;float"/>
                                    </rdf:Description>
                                </owl:someValuesFrom>
                            </owl:Restriction>
                        </owl:intersectionOf>
                    </owl:Class>
                </owl:intersectionOf>
            </owl:Class>
        </owl:equivalentClass>
        <rdfs:subClassOf rdf:resource="&HousesOntology;Building"/>
    </owl:Class>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#HouseExtension -->

    <owl:Class rdf:about="&HousesOntology;HouseExtension">
        <owl:equivalentClass>
            <owl:Class>
                <owl:intersectionOf rdf:parseType="Collection">
                    <owl:Restriction>
                        <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                        <owl:someValuesFrom rdf:resource="&HousesOntology;House"/>
                    </owl:Restriction>
                    <owl:Restriction>
                        <owl:onProperty rdf:resource="&HousesOntology;hasArea"/>
                        <owl:someValuesFrom>
                            <rdf:Description>
```

```
                            <rdf:type rdf:resource="&owl;DataRange"/>
                            <owl11:onDataRange rdf:resource="&xsd;float"/>
                            <owl11:maxInclusive rdf:datatype="&xsd;int">35</owl11:maxInclusive>
                        </rdf:Description>
                    </owl:someValuesFrom>
                </owl:Restriction>
                <rdf:Description rdf:about="&HousesOntology;Building"/>
            </owl:intersectionOf>
        </owl:Class>
    </owl:equivalentClass>
    <rdfs:subClassOf rdf:resource="&HousesOntology;Building"/>
</owl:Class>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#MidTerracedHouse -->

<owl:Class rdf:about="&HousesOntology;MidTerracedHouse">
    <owl:equivalentClass>
        <owl:Class>
            <owl:intersectionOf rdf:parseType="Collection">
                <rdf:Description rdf:about="&HousesOntology;House"/>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl11:onClass rdf:resource="&HousesOntology;House"/>
                    <owl:minCardinality
rdf:datatype="&xsd;nonNegativeInteger">2</owl:minCardinality>
                </owl:Restriction>
            </owl:intersectionOf>
        </owl:Class>
    </owl:equivalentClass>
    <rdfs:subClassOf rdf:resource="&HousesOntology;House"/>
</owl:Class>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#Outbuilding -->

<owl:Class rdf:about="&HousesOntology;Outbuilding">
    <owl:equivalentClass>
        <owl:Class>
            <owl:intersectionOf rdf:parseType="Collection">
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;hasArea"/>
                    <owl:someValuesFrom>
                        <rdf:Description>
                            <rdf:type rdf:resource="&owl;DataRange"/>
                            <owl11:onDataRange rdf:resource="&xsd;float"/>
                            <owl11:maxInclusive rdf:datatype="&xsd;int">35</owl11:maxInclusive>
                        </rdf:Description>
                    </owl:someValuesFrom>
                </owl:Restriction>
                <owl:Class>
                    <owl:complementOf>
                        <owl:Restriction>
                            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                            <owl:someValuesFrom rdf:resource="&HousesOntology;House"/>
                        </owl:Restriction>
                    </owl:complementOf>
```

```
                    </owl:Class>
                    <rdf:Description rdf:about="&HousesOntology;Building"/>
                </owl:intersectionOf>
            </owl:Class>
        </owl:equivalentClass>
        <rdfs:subClassOf rdf:resource="&HousesOntology;Building"/>
    </owl:Class>




    <!-- http://www.semanticweb.org/ontologies/2008/4/HousesOntology-Test3TouchRelation-
terraces.owl#SemiDetachedHouse -->

    <owl:Class rdf:about="&HousesOntology-Test3TouchRelation-terraces;SemiDetachedHouse">
        <owl:equivalentClass>
            <owl:Class>
                <owl:intersectionOf rdf:parseType="Collection">
                    <owl:Class>
                        <owl:complementOf rdf:resource="&HousesOntology;EndTerracedHouse"/>
                    </owl:Class>
                    <owl:Restriction>
                        <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                        <owl11:onClass rdf:resource="&HousesOntology;House"/>
                        <owl:maxCardinality
rdf:datatype="&xsd;nonNegativeInteger">1</owl:maxCardinality>
                    </owl:Restriction>
                    <owl:Class>
                        <owl:complementOf rdf:resource="&HousesOntology;DetachedHouse"/>
                    </owl:Class>
                    <rdf:Description rdf:about="&HousesOntology;House"/>
                </owl:intersectionOf>
            </owl:Class>
        </owl:equivalentClass>
        <rdfs:subClassOf rdf:resource="&HousesOntology;House"/>
    </owl:Class>




    <!-- http://www.w3.org/2002/07/owl#Thing -->

    <owl:Class rdf:about="&owl;Thing"/>




    <!--
    ///////////////////////////////////////////////////////////////////////////////////
    //
    // Individuals
    //
    ///////////////////////////////////////////////////////////////////////////////////
     -->




    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376989 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376989">
        <rdf:type rdf:resource="&HousesOntology;EndTerracedHouse"/>
```

```xml
            <rdf:type rdf:resource="&HousesOntology;House"/>
            <rdf:type>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                    <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
                </owl:Restriction>
            </rdf:type>
            <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376990"/>
            <HousesOntology:hasArea rdf:datatype="&xsd;float">87.6</HousesOntology:hasArea>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376990 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376990">
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">2</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376989"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376991"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">71.5</HousesOntology:hasArea>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376991 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376991">
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">2</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">77.4</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376990"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376992"/>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376992 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376992">
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
```

```
        </rdf:type>
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377010"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376993"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376991"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377011"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">72.5</HousesOntology:hasArea>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376993 -->

    <HousesOntology:MidTerracedHouse rdf:about="&HousesOntology;osgb1000040376993">
        <rdf:type rdf:resource="&HousesOntology;Building"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377010"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376992"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376994"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">72.1</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377011"/>
    </HousesOntology:MidTerracedHouse>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376994 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376994">
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377008"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">73.1</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377009"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376995"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376993"/>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376995 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376995">
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
```

```
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377009"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377008"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">74.6</HousesOntology:hasArea>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376994"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376996"/>
</HousesOntology:Building>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376996 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376996">
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377006"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376997"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376995"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377007"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">71.5</HousesOntology:hasArea>
</HousesOntology:Building>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376997 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376997">
    <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376998"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377007"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">73.8</HousesOntology:hasArea>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377006"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376996"/>
</HousesOntology:Building>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376998 -->

<HousesOntology:MidTerracedHouse rdf:about="&HousesOntology;osgb1000040376998">
    <rdf:type rdf:resource="&HousesOntology;Building"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
```

```
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376997"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377004"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376999"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">73.1</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377005"/>
    </HousesOntology:MidTerracedHouse>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040376999 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040376999">
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">4</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376998"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377005"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">73.3</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377004"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377000"/>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377000 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377000">
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376999"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377001"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">75.1</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377003"/>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377001 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377001">
        <rdf:type rdf:resource="&HousesOntology;MidTerracedHouse"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
```

```
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377002"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377000"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">68.9</HousesOntology:hasArea>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377003"/>
</HousesOntology:Building>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377002 -->

<HousesOntology:EndTerracedHouse rdf:about="&HousesOntology;osgb1000040377002">
    <rdf:type rdf:resource="&HousesOntology;House"/>
    <rdf:type rdf:resource="&HousesOntology;Building"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377001"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">79</HousesOntology:hasArea>
</HousesOntology:EndTerracedHouse>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377003 -->

<HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377003">
    <rdf:type rdf:resource="&HousesOntology;Building"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">2</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377001"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">3.4</HousesOntology:hasArea>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377000"/>
</HousesOntology:HouseExtension>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377004 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377004">
    <rdf:type rdf:resource="&HousesOntology;HouseExtension"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
```

```xml
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.5</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377005"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376998"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376999"/>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377005 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377005">
        <rdf:type rdf:resource="&HousesOntology;HouseExtension"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.6</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376999"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376998"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377004"/>
    </HousesOntology:Building>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377006 -->

    <HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377006">
        <rdf:type rdf:resource="&HousesOntology;Building"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376996"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377007"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376997"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.2</HousesOntology:hasArea>
    </HousesOntology:HouseExtension>



    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377007 -->

    <HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377007">
        <rdf:type rdf:resource="&HousesOntology;Building"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377006"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376997"/>
```

```
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376996"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.1</HousesOntology:hasArea>
    </HousesOntology:HouseExtension>




    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377008 -->

    <HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377008">
        <rdf:type rdf:resource="&HousesOntology;HouseExtension"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376994"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376995"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.8</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377009"/>
    </HousesOntology:Building>




    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377009 -->

    <HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377009">
        <rdf:type rdf:resource="&HousesOntology;Building"/>
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.1</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376995"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377008"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376994"/>
    </HousesOntology:HouseExtension>




    <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377010 -->

    <HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377010">
        <rdf:type>
            <owl:Restriction>
                <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
            </owl:Restriction>
        </rdf:type>
        <rdf:type rdf:resource="&HousesOntology;Building"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376992"/>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376993"/>
        <HousesOntology:hasArea rdf:datatype="&xsd;float">3.5</HousesOntology:hasArea>
        <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377011"/>
```

```
</HousesOntology:HouseExtension>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377011 -->

<HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377011">
    <rdf:type rdf:resource="&HousesOntology;Building"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">3</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">3</HousesOntology:hasArea>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377010"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376993"/>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040376992"/>
</HousesOntology:HouseExtension>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377012 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377012">
    <rdf:type>
        <owl:Class>
            <owl:complementOf>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl:someValuesFrom rdf:resource="&HousesOntology;Building"/>
                </owl:Restriction>
            </owl:complementOf>
        </owl:Class>
    </rdf:type>
    <rdf:type rdf:resource="&HousesOntology;Outbuilding"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">22.3</HousesOntology:hasArea>
</HousesOntology:Building>




<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377014 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377014">
    <rdf:type>
        <owl:Class>
            <owl:complementOf>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl:someValuesFrom rdf:resource="&HousesOntology;Building"/>
                </owl:Restriction>
            </owl:complementOf>
        </owl:Class>
    </rdf:type>
    <rdf:type rdf:resource="&HousesOntology;Outbuilding"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">12.9</HousesOntology:hasArea>
</HousesOntology:Building>
```

```
<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377732 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377732">
    <rdf:type rdf:resource="&HousesOntology;DetachedHouse"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377754"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">119.9</HousesOntology:hasArea>
</HousesOntology:Building>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377733 -->

<HousesOntology:Building rdf:about="&HousesOntology;osgb1000040377733">
    <rdf:type>
        <owl:Class>
            <owl:complementOf>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl:someValuesFrom rdf:resource="&HousesOntology;Building"/>
                </owl:Restriction>
            </owl:complementOf>
        </owl:Class>
    </rdf:type>
    <rdf:type rdf:resource="&HousesOntology;DetachedHouse"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">103.4</HousesOntology:hasArea>
</HousesOntology:Building>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377754 -->

<HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377754">
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&HousesOntology;touches"/>
            <owl11:onClass rdf:resource="&HousesOntology;Building"/>
            <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
        </owl:Restriction>
    </rdf:type>
    <rdf:type rdf:resource="&HousesOntology;Building"/>
    <HousesOntology:hasArea rdf:datatype="&xsd;float">4.5</HousesOntology:hasArea>
    <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377732"/>
</HousesOntology:HouseExtension>



<!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377759 -->

<HousesOntology-Test3TouchRelation-terraces:SemiDetachedHouse
rdf:about="&HousesOntology;osgb1000040377759">
```

```
                <rdf:type rdf:resource="&HousesOntology;Building"/>
                <rdf:type>
                    <owl:Restriction>
                        <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                        <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                        <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">2</owl:cardinality>
                    </owl:Restriction>
                </rdf:type>
                <rdf:type rdf:resource="&HousesOntology;House"/>
                <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377769"/>
                <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377760"/>
                <HousesOntology:hasArea rdf:datatype="&xsd;float">98.6</HousesOntology:hasArea>
        </HousesOntology-Test3TouchRelation-terraces:SemiDetachedHouse>



        <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377760 -->

        <HousesOntology:House rdf:about="&HousesOntology;osgb1000040377760">
            <rdf:type>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                    <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
                </owl:Restriction>
            </rdf:type>
            <rdf:type rdf:resource="&HousesOntology-Test3TouchRelation-terraces;SemiDetachedHouse"/>
            <rdf:type rdf:resource="&HousesOntology;Building"/>
            <HousesOntology:hasArea rdf:datatype="&xsd;float">97.7</HousesOntology:hasArea>
            <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377759"/>
        </HousesOntology:House>



        <!-- http://www.semanticweb.org/ontologies/2008/3/HousesOntology.owl#osgb1000040377769 -->

        <HousesOntology:HouseExtension rdf:about="&HousesOntology;osgb1000040377769">
            <rdf:type rdf:resource="&HousesOntology;Building"/>
            <rdf:type>
                <owl:Restriction>
                    <owl:onProperty rdf:resource="&HousesOntology;touches"/>
                    <owl11:onClass rdf:resource="&HousesOntology;Building"/>
                    <owl:cardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:cardinality>
                </owl:Restriction>
            </rdf:type>
            <HousesOntology:touches rdf:resource="&HousesOntology;osgb1000040377759"/>
            <HousesOntology:hasArea rdf:datatype="&xsd;float">14.7</HousesOntology:hasArea>
        </HousesOntology:HouseExtension>
    </rdf:RDF>
```

Example OWL syntax for defined Urban Block individuals:

Below is the OWL syntax for one defined urban block individual that is used for the inference of types of blocks and districts. An individual's syntax can be very long because some of the blocks contain hundreds of building individuals.

```xml
<!-- http://www.semanticweb.org/ontologies/2008/5/GlasgowSample_HouseTypes.owl#UB63991 -->

<owl:Thing rdf:about="&GlasgowSample_HouseTypes;UB63991">
    <rdf:type rdf:resource="&GlasgowSample_HouseTypes;BlockMixedHouses"/>
    <rdf:type rdf:resource="&GlasgowSample_HouseTypes;DistrictMixedHouses"/>
    <rdf:type>
        <owl:Restriction>
            <owl:onProperty rdf:resource="&GlasgowSample_HouseTypes;contains"/>
            <owl:allValuesFrom>
                <owl:Class>
                    <owl:oneOf rdf:parseType="Collection">
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377135"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377166"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377132"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377161"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377143"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377137"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377136"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377164"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377138"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377162"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377130"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377169"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377165"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377172"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377167"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377133"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377173"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377142"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377159"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377129"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377134"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377170"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377131"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377168"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377140"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377141"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377139"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377171"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377163"/>
                        <rdf:Description rdf:about="&GlasgowSample;osgb1000040377160"/>
                    </owl:oneOf>
                </owl:Class>
            </owl:allValuesFrom>
        </owl:Restriction>
    </rdf:type>
    <GlasgowSample_HouseTypes:hasPercentageDetached
rdf:datatype="&xsd;float">0</GlasgowSample_HouseTypes:hasPercentageDetached>
    <GlasgowSample_HouseTypes:hasPercentageTerraces
rdf:datatype="&xsd;float">20</GlasgowSample_HouseTypes:hasPercentageTerraces>
    <GlasgowSample_HouseTypes:hasPercentageSemis
rdf:datatype="&xsd;float">80</GlasgowSample_HouseTypes:hasPercentageSemis>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377129"/>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377130"/>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377131"/>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377132"/>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377133"/>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377134"/>
    <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377135"/>
```

```
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377136"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377137"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377138"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377139"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377140"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377141"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377142"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377143"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377159"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377160"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377161"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377162"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377163"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377164"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377165"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377166"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377167"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377168"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377169"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377170"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377171"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377172"/>
        <GlasgowSample_HouseTypes:contains rdf:resource="&GlasgowSample;osgb1000040377173"/>
        <GlasgowSample_HouseTypes:connectedTo rdf:resource="&GlasgowSample_HouseTypes;UB63833"/>
        <GlasgowSample_HouseTypes:connectedTo rdf:resource="&GlasgowSample_HouseTypes;UB64608"/>
        <GlasgowSample_HouseTypes:connectedTo rdf:resource="&GlasgowSample_HouseTypes;UB64610"/>
        <GlasgowSample_HouseTypes:connectedTo rdf:resource="&GlasgowSample_HouseTypes;UB66001"/>
        <GlasgowSample_HouseTypes:connectedTo rdf:resource="&GlasgowSample_HouseTypes;UB66244"/>
    </owl:Thing>
```