A Methodology for Inferring Higher Level Semantic Information from Spatial Databases

Marie-Kristina Thomson and Roderic Béra

Department of Civil, Environmental & Geomatic Engineering, University College London,
Gower Street, London, WC1E 6BT
Tel. +44 (0) 20 7679 2741
{thomson, rbera}@ge.ucl.ac.uk

KEYWORDS: ontology, knowledge representation, interpretation, land use, mapping

1. Introduction

The lack of high level semantic descriptions within geospatial data is a major critique point in regards to their handling and usage (Freksa and Barkowsky, 1995; Peuquet, 1988; Burrough and Frank, 1995; Frank, 1992). Topographic data are a striking example, as the only explicit knowledge they contain is positional. They can reveal a lot more, however, as spatial configuration are the result of functional and social processes (Lévi-Strauss, 1963). In this context the present research seeks to expose implicit information within a database to enrich its contents by generating a description that assigns functional meanings (i.e. higher level semantic information) to the map primitives. Such an approach is expected to ultimately improve the responding of data providers to customer needs and specific applications (Lüscher *et al.*, 2007).

The following sections outline how the use of knowledge representation for the explicit formalisation of knowledge about object constellations existent in a topographic map can be exploited. Section 2 discusses the pertinence of a distinction between data and semantics and how to link both aspects, section 3 proposes a model framework for the extraction of higher level knowledge and exemplifies the approach on the case of the notion of residential area, before drawing some temporary conclusions (section 4).

2. Decomposing the Mapping Problem between Knowledge and Geographic Data

Ontological approaches emerge as an alternative representation for objects stored in a geo-spatial database, thus providing new opportunities for intelligent information processing (Torres *et al.*, 2005; Hart, 2007; Hereth *et al.*, 2000; Klien and Lutz, 2005). An ontology appears as a mediating instance between the captured reality (spatial data), and higher level knowledge as perceived and conceptualised by humans. People can readily assign such knowledge to geographic data, as shown in Thomson and Béra (2007) where respondents interpreted topographic maps for land use information. The visual recognition of a 'residential' pattern on a map, for example, is characterised by the ability to interpret and categorise new observation data (Mennis *et al.,* 2000). By using existing knowledge about the objects' spatial configuration and distinctive spatial properties (such as size, shape, layout, and other criteria), a person can identify, to a certain extent, areas in a map by their primary purpose. This holds especially for the various types of residential areas (Thomson and Béra, 2007).

In order to translate this ability into mechanised ways, 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). In order to capture the essence of human perception, understanding, reasoning and intuition, a machine-compatible approach needs to resemble the levels of organisation in general intelligence, as summarised by Yudkowsky (2007). This organisation can be adopted to reflect the processing of geographic data in an intelligent manner. For instance, on a set of spatial structures, such as topographic features, reasoning processes operate based on similarity judgement, proximities, shapes, sizes, etc. (Thomson and Béra, 2007), in ways similar to those investigated by Gestalt psychology. Concepts can be associated to these structures and processes to form a knowledge representation. Higher level concepts can then be built incrementally from these re-usable, primitive concepts as disposable one-time conceptual entities dependent on the information that is required by the user. Inference is achieved by classifying instances from the underlying database in terms of constrained object configurations and other concept parts as ascribed by the high level structures.

Such an approach can be termed semantic data processing since inference is based on reasoning with concepts about a data's semantics (figure 1). Similarities can be drawn to natural language processing (e.g. Katz and Fodor, 1963; Lewis, 1970; Blutner, 2002), where (a) morphological knowledge relates to the understanding of multiple forms of objects and their spatial order, (b) syntactic knowledge states structural information and relationships about the order of objects in a group signalling a particular meaning, and (c) semantic knowledge defines meaning about context and how concepts relate to objects and relations in the world. The question is how to get from raw input, i.e. spatial data, to meaning, and from meanings to customised output (ultimately new map products).

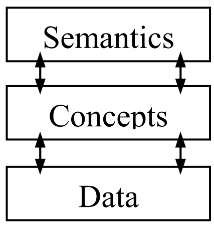

Figure 1. Semantic Data Processing

Data representation is treated 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). Richer knowledge is derived from meaningful object configurations, special relations, perception (Gestalt principles) and context, which needs to be grounded in the data to allow a one-to-one mapping. In a domain which manifests itself through its underlying geographic reality, such a grounding can be achieved by decomposing the rich knowledge to its finest level of detail in terms of its 'syntax', i.e. information about the structure of spatial objects (such as roads, buildings, land, and water), and how these comprise larger units that convey functional meaning, feature attributes, and spatial relations. This is exemplified in the conceptual decomposition of 'residential area' in figure 2. 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.

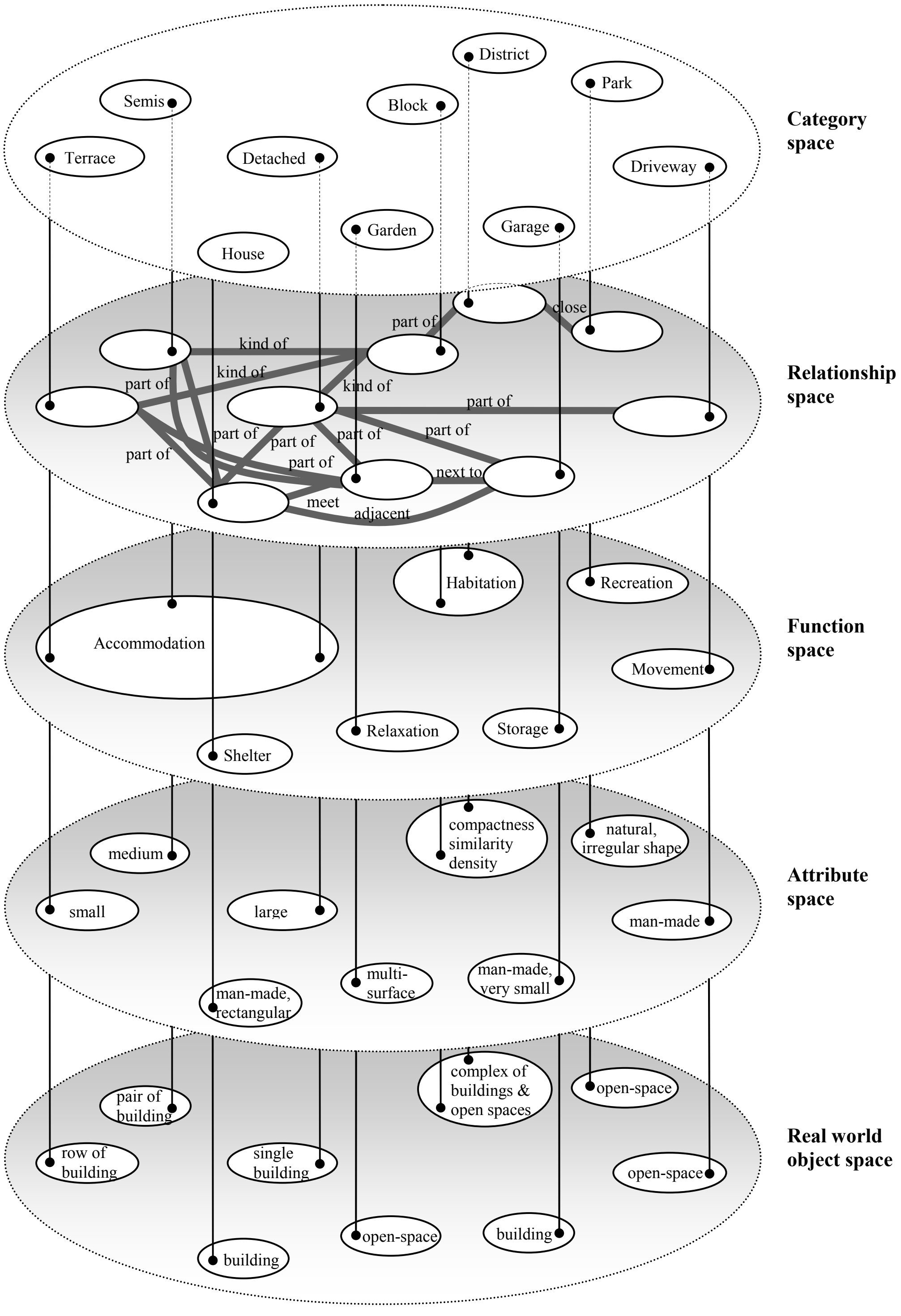

Figure 2. Decomposing the high level concept 'residential area' into its parts (derived from Thomson and Béra, 2007)

3. Model-Based Recognition of the Concept 'Residential Area' from Topographic Data

The outlined mapping in figure 2 translates into a conceptual framework that allows the processing and assignment of higher level semantic information to spatial data. Figure 3 illustrates how implicitly represented knowledge within a spatial database (such as residential area) can be inferred from knowledge that is explicitly defined through ontologies.

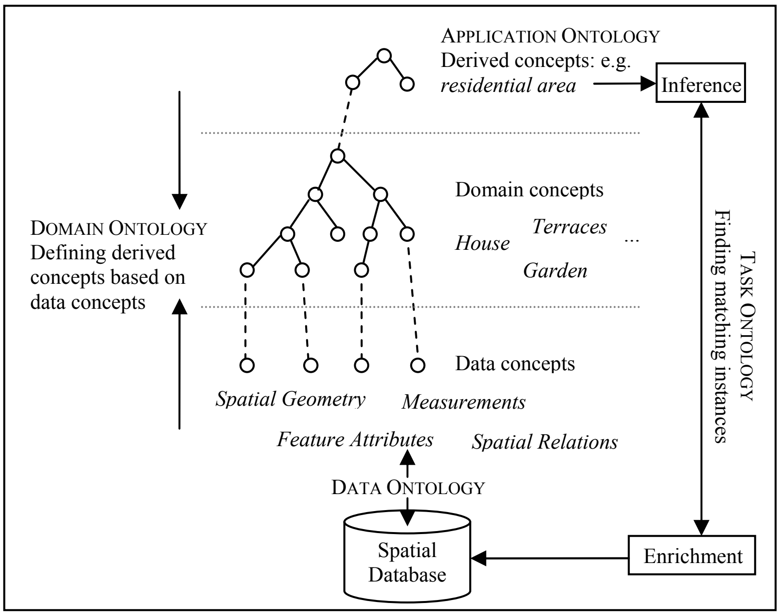

Figure 3. Relating higher level concepts to the data structure

The modelling takes a bottom-up, 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. Top-down modelling is adopted to define the part-whole aggregation based on certain criteria and rules which will constrain reasoning of the description logic. The incremental steps and different aggregate levels are illustrated in figure 4.

This methodology uses description logic (in a way similar to Neumann and Möller (2006) 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 set 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 which decompose into further subsystems arranged into distinct hierarchies of taxonomies and partonomies. Although we have to be careful not to force the diversity of functions into a narrow range of concepts, providing a system structure enables inference of higher level information based on reasoning about the defined concepts and finding its relevant instances. Therefore, 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.

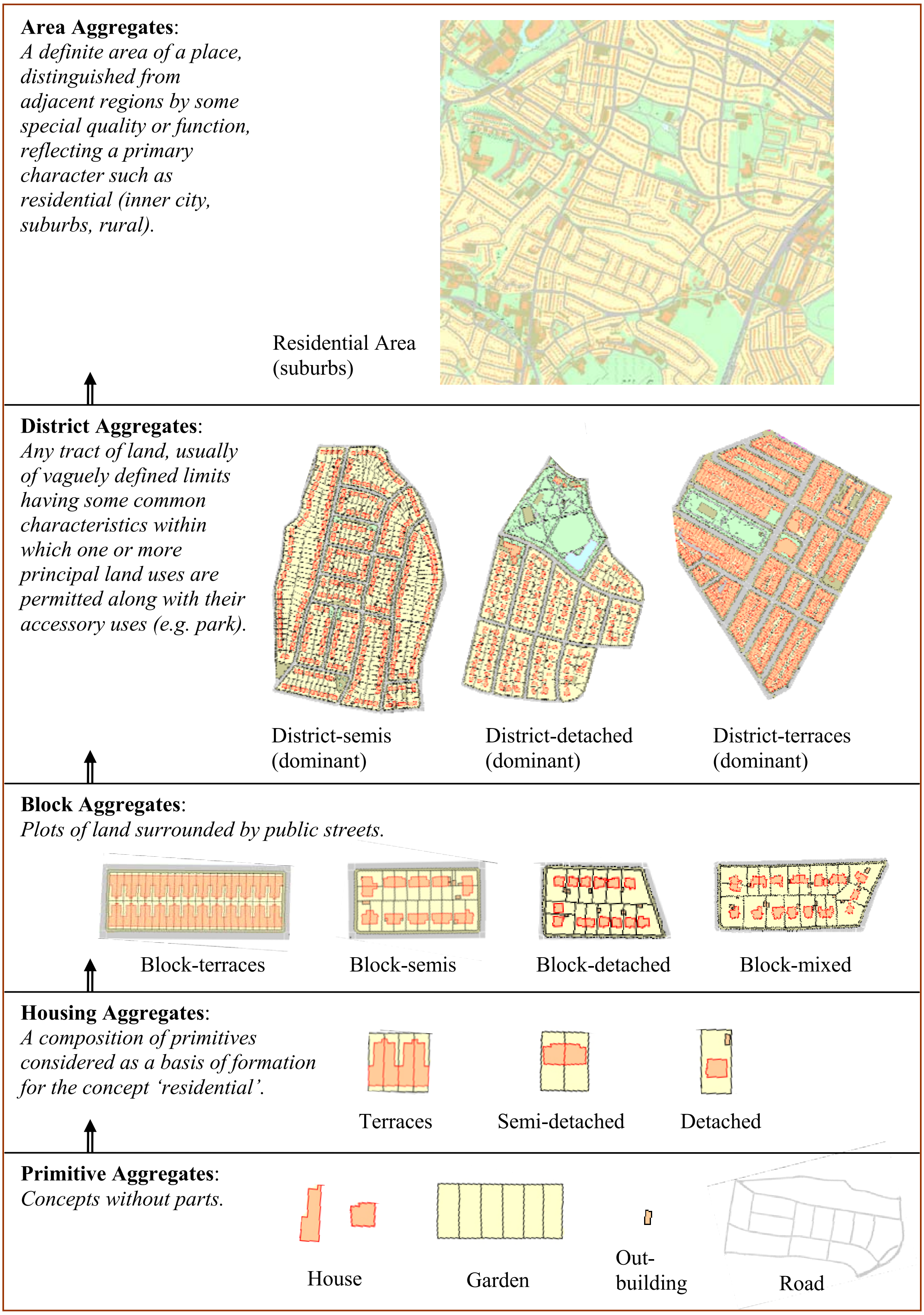

Figure 4. A system structure of aggregate concepts defining compositional elements of 'residential area'

4. Conclusion

This paper reviewed part-whole aggregation in the context of deriving higher level concepts from spatial data. By demonstrating how a hierarchy of spatial composites conveys implicit functional meaning, similar to the syntax of a natural language sentence, we can conclude that such meaning can be made explicit and derived from spatial data. The presented methodology is important for making data more definitive, intelligent and accessible. Function in particular has been identified as one of the five basic ontological relations that make geographic information more explicitly meaningful (Kuhn, 2007). The enrichment of spatial databases with such meaningful information is an essential part of the abstraction process for producing on-demand, customised, and multi-representational map products (Regnauld, 2007). Not only is there the need to reflect more the way people perceive the world (Mennis *et al.*, 2000), but to develop new frameworks that maintain a direct link to the data but can also be disassociated to ensure flexibility and allow the association with different data sources. Future work concerns the implementation of the proposed system through current technologies evolving around ontologies and description logics.

5. Acknowledgements

This work is part of a project funded by the Engineering and Physical Sciences Research Council as well as Ordnance Survey, the national mapping agency of Great Britain. Many thanks go to Nicolas Regnauld (OS) and Muki Haklay (UCL) for providing valuable comments to the writing of this extended abstract.

References

Barr, S. and Barnsley, M. (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

Batty, M. and Longley, P. (1994) *Fractal Cities*. London: Academic Press Limited

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

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

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

Freksa, C. and Barkowsky, T. (1995) On the Relation between Spatial Concepts and Geographic Objects. In Burrough, P. and Frank, A. (Eds) *Geographic Objects with Undetermined Boundaries*. London: Taylor & Francis

Haggett, P. and Chorley, R. J. (1969) *Network analysis in geography*. London: Edward Arnold

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

Hereth, J., Stumme, G., Wille, R. and Wille, U. (2000) Conceptual knowledge discovery and data analysis. In Mineau, G. and Ganter, B. (Eds): *International Conference on Conceptual Structures*, LNCS 1867, pages 421-437. Springer-Verlag

Katz, J. J. and Fodor, J. A. (1963) The structure of a semantic theory. *Language* **39**(2): 170-210

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

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 Berlin Heidelberg
Lévi-Strauss, C. (1963) *Structural Anthropology*. New York: Basic Books
Lewis, D. (1970) General semantics. *Synthese* **22**(1-2): 18-67
Lüscher, P., Burghardt, D. and Weibel, R. (2007) Ontology-driven enrichment of spatial databases. 10th ICA Workshop on Generalisation and Multiple Representation, 2-3 August 2007, Moscow
Marshall, S. (2005) *Streets and Patterns*. London: Spon Press
Mennis, J.L., Peuquet, D.J. and Qian, L. (2000) A conceptual framework for incorporating cognitive principles into geographical database representation. *International Journal of Geographical Information Science* **14**(6): 501-520
Neumann, B. and Möller, R. (2006) On Scene Interpretation with Description Logics. In Christensen, H. I. and Nagel, H.-H. (Eds.): Cognitive Vision Systems. LNCS 3948, pp.247-275 Springer-Verlag Berlin Heidelberg
Regnauld, N. (2007) A distributed system architecture to provide on-demand mapping. *Proceedings of the XXIII International Cartographic Conference*, 4-10 August 2007, Moscow, Russia
Peuquet, D.J. (1988) Representations of Geographic Space: Toward a Conceptual Synthesis. *Annals of the Association of American Geographers* **78**(3): 375-394
Thomson, M.-K. and Béra, R. (2007) 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
Torres, M., Quintero, R., Moreno, M. and Fonseca, F. (2005) Ontology-driven description of spatial data for their semantic processing. In Rodríguez *et al*. (Eds.): *GeoS 2005*, LNCS 3799, pp. 242-249, Springer-Verlag Berlin Heidelberg
Yudkowsky, E. (2007) Levels of Organization in General Intelligence. In Goertzel, B. and Pennachin, C. (Eds.): Artificial General Intelligence, pp.389-501 Springer-Verlag Berlin Heidelberg

Biography

Marie-Kristina Thomson is a PhD student in the department of Civil, Environmental and Geomatic Engineering at UCL researching ways to infer higher level semantic information from spatial databases. She holds a BSc (Hons) in Remote Sensing and GIS from Bath Spa University, and she is currently student representative of the Remote Sensing and Photogrammetry Society.

Dr. Roderic Béra is lecturer in geographical information science at the UCL department of Civil, Environmental and Geomatic Engineering (CEGE). His research interests lie mainly in ontology for modelling relative proximities, proximity and space syntax, and the influence of geography in web-based social networks. He was previously a teaching and research assistant at the French Naval Academy Research Institute.