MANSION-GS: seMANtics as the n-th dimenSION for Geographic Space

Salvatore F. Pileggi
The University of Auckland
f.pileggi@auckland.ac.nz

Robert Amor
The University of Auckland
trebor@cs.auckland.ac.nz

Abstract
The extended understanding of geographic ecosystems, including the physical and logical description of space with associated data and activities as well as the dynamics inside, poses complex scenarios that cannot be reflected on a simple geographic-oriented data model. The main purpose of this current work is the conceptual integration of a physical space model with dynamic logic support able to describe the relations amongst the different elements composing the space as well as the relations between spaces and external elements. In the context of this work, semantics have the critical and central role of connecting and relating the different dimensions on the space, even though they are mostly a virtual dimension in the overall model.

Keywords
Semantic Technologies, Knowledge Modelling, Ontology Development and GIS.

MANSION-GS: seMANtics as the n-th dimenSION for Geographic Space

1. Introduction

The definition of semantics associated with a space in order to describe the physical and logic composition of the space itself as well as the relations among the different components is a popular topic for various domains and applications. If the understanding of the geographic space is extended to geographic ecosystems resulting from the combination of the physical space, associated data and activities and dynamics inside them, then the complexity of target semantics considerably increases impacting the integration of common models with abstracted logic views to support the complexity of the reality.

Those needs seem to be validated simply by analysing normal geographic data or data associated with geographic spaces. In fact, a large number of data sources are currently available. They have different foci and are normally used in different contexts for different purposes. Those heterogeneous data provide multiple perspectives of the same physical environment and, so, they cannot be considered as logically independent (stand-alone) information, as a first analysis would suggest. Furthermore, the same limitations about space representation affect most distributed data represented by poor data models that do not allow complex relations among data and so cannot provide any support for data abstraction and knowledge building. In this context, the only possible relation between data and space is offered by a fundamental geographic-based focus. The overall scenario could be summarized as an environment full of information (gross data) but poor in terms of knowledge and intelligibility.

If geographic spaces and available data are structured and modelled with rich semantic models (Pileggi and Amor. 2013) and data are considered together (both with contextual information), not as a simple federation but as part of a common semantic ecosystem, the capabilities in terms of analysis increase strongly. Those capabilities are the key and critical issue for complex applications aimed at building some kind of knowledge on the base of gross data.

Another limiting aspect of knowledge building on geographic data is the complete lack of social focus. Public spaces are living environments in which people develop their everyday life or have concrete experiences (e.g. tourism activities). If the data analysis just refers to static data, the results have a static understanding, with analysis capabilities strongly limited in practice (even if they are not in theory). Most of the knowledge of a public space is “inside” people. As individuals of course, but also (and above all) as part of communities. Integrating data about the city with related social information enables a consistent and rich social semantic ecosystem around spaces. The point of convergence for these trends is mostly represented by semantic technologies that allow rich and potentially interoperable data models for heterogeneous complex data.

The main purpose of this current work is the conceptual integration of a physical space model with a dynamic logic support capability able to describe the relations amongst the different elements composing the space as well as the relations between spaces and external elements.

The most direct consequence of a rich description of the space is the need to bring together different heterogeneous layers of knowledge as dimensions on the space. Resulting models are hard to be managed and mostly useless if they are not defined according to formal schemas in a context of high expressivity and interoperability. That is mostly the role of semantic technologies. So, in the context of this work, semantics have the critical and central role of connecting and relating the different dimensions on the space, even though they are a virtual dimension more in the overall model. The methodology adopted for domain analysis and the consequent ontology development is a strong simplification of the model proposed in (Pileggi,Calvo-Gallego and Amor. 2013). In the context of this work, we exclusively focus on data (and not on data and services as in (Pileggi,Calvo-Gallego and Amor. 2013)). Furthermore, the nature and class of domain resource is mostly known a priori. The main resources currently considered are the spaces, as well as related data and social objects (Pileggi et al. 2012). Application specific concepts are included in the model but their description is out of the paper scope. The section on related work closes the introductive part of the paper. The core part follows it and is composed of three different subsections which deal with the main concept (MANSION), its implementation (MANSION ontology) and a couple of application scenarios. As usual, the paper finishes with a section summarizing the provided contribution.

2. Related Work

The general need for semantics inside complex systems is a common concern and, of course, it affects the space domain as well as any kind of application involving logic for geographic environments. The problem of missing semantics is more problematic when heterogeneous data (Wiegand et al. 2003) are involved. Applications are forced to work on simple data models in a context of ambiguity and a lack of interoperability. The high-level analysis of human environments based on heterogeneous data (Gao et al. 2010, Yue et al. 2011) is progressively increasing in importance in the context of different domains and applications (such as urban planning (Gomes et al. 2012, Zhang et al. 2010). Most recent applications propose approaches based on rich data models (NG et al. 2012, Strintzis et al. 2009), such as ontological models (Mao and Li. 2011). Machine-understandable knowledge mapped on rich data models can be obtained by applying techniques, such as rule-based extraction (Zhang et al. 2009) or image processing (Ardizzone et al. 2012). Most semantic environments currently available have a clear application-specific focus (e.g. (Luckel et al. 2009, Shahabi et al. 2010, Eldien et al. 2009, Zheng-yu et al. 2009). Furthermore, even though the latest trends assume a certain participation of people (Muramoto et al. 2006) or models based on human-behaviour (Suhong et al. 2010), the progressive socialization of IT and its intrinsic capabilities are not properly reflected on data models and related applications (Yue et al. 2011).

The ontological approach proposed in the paper would overcome limitations previously described by:

- Providing a multi-layer semantic support framework which is built on the top of a generic base layer and eventually integrated with domain specific layers.
- Integrating the physical description of the space with a logic-based understanding of it. This is partially or completely missed in most works (e.g. Janowicz et al. 2013 and Koubarakis et al. 2012).
- Focusing on relations. Relations among space components are not limited to physical relations. They are modelled according to an open and extensible understanding of the

vocabularies that allows domain/application specific relations in the context of a generic model.

- The same open approach used to describe the space is followed concerning data.
- Explicit management of Social Objects (Pileggi at al. 2012) in order to facilitate the representation and understanding of social data as well as their relations to the space.

3. MANSION Concept

Semantic support (seMANtics as the n-th dimenSION, MANSION) for the problems previously introduced is composed of three independent, even if logically related, layers (Figure 1, left). They are designed and developed according to a bottom-up approach:

(1) *Support for Geo-Resources and related Data.* This is the base layer on top of which the other layers are logically built. It is a completely application independent layer aimed at the representation of the space as a semantic geo-resource. Apart from the representation of the space itself, it also provides a semantic view of the space with the set of relations involving the different elements. Semantic features and data layers are integrated in order to assure a semantic consistency to the whole model.

(2) *Support for Social Resources.* As previously mentioned in the introductory part of the paper, social trends are strongly affecting global information and the understanding of the information itself. Social data are for instance more complex than any other kind of data. And, at the same time, they can provide added value as they are rich in terms of implicit and explicit knowledge (Pileggi at al. 2012). That is the reason for providing the model with a specific extension in order to relate social data flows to the space exactly as for one dashboard data. The models for representing social objects are complex and are not the objective of this work.

(3) *High-level Semantics.* These are domain/application specific layers that introduce application specific concepts as well as the relations between these concepts and lower layers. This is normally a multi-layer data structure. Details about application specific layers are very interesting but out of the scope of this paper.

The MANSION model has a general focus. In this paper we consider the Geographic Space as a target domain. The specialization of the model to the Geographic Space is referred as MANSION-GS. Focusing on the base layer of the MANSION semantic support, the following subsections address the Space Model adopted, the description of main vocabularies provided, and the model of layers and features currently in use.

3.1 Space Model

The space model is one of the key factors for the effectiveness of the whole model. It should be expressive enough to represent complex relations among the elements composing the space but, at the same time, simple enough to be understood and used in practice. The model adopted in MANSION (Figure 1, right) is explicitly designed according to a perspective-oriented view of the information. In order to be recognized as geo-resource an element has to be defined as a *Container*. This is a dynamic concept and different kinds of containers can be defined (e.g. geo-containers and logic containers). Details about the containers are provided below. Each container has, for instance, a *logic level* associated to it. In most cases it is associated to the level of abstraction or to some physical parameter. It is normally assumed higher (or more abstracted) containers have a higher value of the parameter. The parameter's value can be explicitly defined or inferred from semantic properties considering the space as a whole. Focusing on a concrete container (called *Main Container*), the model provides the semantic view of the space from the Main Container's perspective as well as the relation to the context information. Each container can be composed of other containers or it can include elements or containers even if they are not strictly composing it. Normally both composing and included elements should have a logic level equal or less than the container they are in. If not, semantic inconsistences could be introduced. Concerning the inclusion of containers in others, in most cases this is a physical understanding of inclusion. However, specific applications could define some kind of logic not matching the physical understanding. The model fully supports multiple understandings. Furthermore, as in any relation (see next section), the model allows open relations and so inclusion or composition of spaces can be semantically extended in order to specify concrete information or roles. As a container can be related to a lower container, it can also be related to one or more upper containers (*Parent Container* in the model). The parent relationship normally involves a container of equal or higher level but it is conceptually different from previously discussed relations (inclusion and composition). Indeed, the fact that the container *a* is included in container *b* does not imply *b* is the parent of *a*. They are different relations, used in practice for different purposes. Normally, a container inherits properties/data from the parent or provides the parent with properties/data. In a context of ecosystems a container normally does not exist by itself but has (or it is part of) a context. This context is just partially defined by parents and included containers. There may be containers, outside of any logic of composition or property/data flow, that are logically related (*Related Container* in the model). Generic relations particularized by the use of open vocabularies (see next section) involve mostly containers of the same level but there are no semantic inconsistences in relations among containers from different levels.

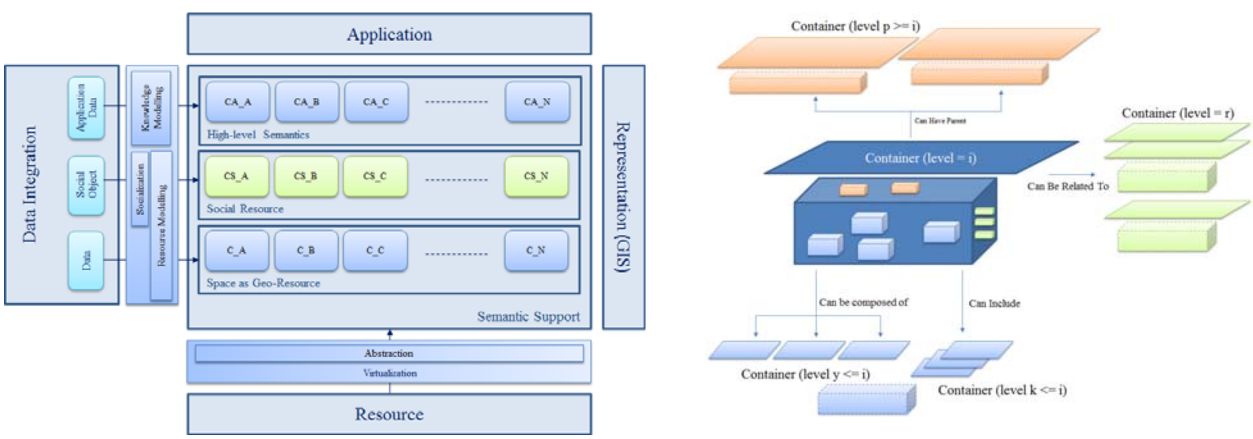

Figure 1: A layered view of MANSION's sematic support (left) and the Space Model in MANSION (right).

3.2 Vocabularies

The MANSION model is expressed as an ontological model. It is common to assume a direct correspondence from ontology to vocabulary. In MANSION each ontology reflects a layer (or a part of one) in the model and these can be more than one vocabulary. That allows us to dynamically classify the concepts as well as providing an understanding of the "public" information, hiding details and concepts that have an internal meaning more than a user level understanding. The logic partitioning of ontologies in vocabularies is strongly dynamic. It can be performed at concept level (if a class is part of a given vocabulary, then each individual of the class is associated to that vocabulary) or at individual level (each individual is directly associated to a vocabulary regardless by its class).

Focusing on the base level of the model, we are currently providing three different vocabularies:

(1) *Domain Vocabulary*. This reflects the space model as described in the previous subsections and a set of concept to classify containers (e.g. country, city, district, suburb, place, building, etc.).

(2) *Input Vocabulary*. This includes any input data (data, features and level will be addressed in the following subsection) aside from the domain.

(3) *Internal Vocabulary*. Low level concepts, concepts for internal use only, or non user level concepts.

3.3 Layers and Features

The ideal complement for the space domain is the direct or indirect association of related data. Data is for instance an application specific concept as well as the semantic features associated to elements of the space and to the related information. Therefore, it is not easy to propose a unique model assuring an open high flexible understanding of features and information. That is normally the main object of the Input Vocabulary as previously defined.

In the MANSION model, the central concept in the input vocabulary is the Feature. A feature can have different types in function of purpose (e.g. classification) and applicable domain (e.g. container or data). Some details are provided in the next section.

The centrality of semantic features is motivated by the fact that any data or semantic layer is defined according to one or more features. Features can be involved in the definition of relations among the space elements (specifying in practice the kind of relations), to group data (semantic data layers) and for any kind of semantic specification inside the model.

4. MANSION-GS Ontology

The main purpose of this section is to provide a descriptive overview of the MANSION-GS Ontology (Figure 2) that implements the MANSION concept for geographic space. Only key high-level concepts will be considered in order to assure an understanding of the model without going through model details. The ontology (Figure 2) is implemented with OWL technology (OWL-Web Ontology Language) as a set of Concepts, Object Properties, Data Properties and Individuals.

The Domain Vocabulary (as described in the previous section) is defined on top of two key concepts:

- *Ecosystem*. It is the "highest" concept the ontology currently recognizes. It has to provide an understanding of what kind of environment is being defined. At the moment, three different ecosystems can be specified: Metropolitan Ecosystem, Ocean Ecosystem (E) and Global Ecosystem. In MANSION, an Ecosystem is explicitly defined by the set of associated semantic features (F) and is implicitly defined by the containers that compose it (eq.1).

$$E_i(F,\_) \leftarrow \cup_i [C(F_i,\_)] \qquad \text{(eq.1)}$$

- *Container.* Regardless of the kind of ecosystem, the space is always represented as a composition of containers with a dynamic set of semantic features. A container can be a Physical Container (if it is reflecting a physical reality, such as a District or a Place) or a Logic Container (if it is dynamically composed of elements not fully matching a physical environment). The difference between those two macro concepts is not well defined and boundaries between them can vary as a function of the application context. For instance, a logic container should include elements that have an intrinsic logic relation even they do not have a strict geographic relation. For example, universities can be composed of different campuses that do not always have a geographic continuity. So, the (overall) university can be defined as a logic container composed of different geographically distributed spaces (campuses). In MANSION, a logic or physical container is defined by a set of semantic features and a set of semantic relations (eq.2).

$$C_i(P/L,\_) \rightarrow [F_i, R_i,\_] \qquad \text{(eq.2)}$$

The Input Vocabulary (as in the previous section) currently includes three classes of concepts:

- *Feature.* This is a set of semantic features that are used to classify or characterize elements of the ontology. They are part of the Input Vocabulary since they have an application specific character and, so, it is natural to manage them as they are input data provided by applications. Apart from a set of Internal Features (that belong the Internal Vocabulary), the ontology currently includes features for the classification of spaces, for the definition of data layers and for the specification of areas with special features (e.g. protected areas).
- *Data.* As in the common meaning.
- *Social Object.* These are defined as in (Pileggi at al. 2012) and they are, for instance, the most complex data considered in the model. Their structure could be considered as an application independent topic but, in practice, due to their intrinsic complexity they are application specific elements.

Internal classes of concepts of interest are:

- *Filter.* This is a set of inferred concepts that improve the usability of the ontology limiting the complexity of the queries.
- *Relation.* One of the key features of MANSION is the capability to provide open relations between the elements composing the model. Relations can be uni-directional or bi-directional relations and they include a set of Object Properties that define the nature of the relation as well as further elements involved in the relation.

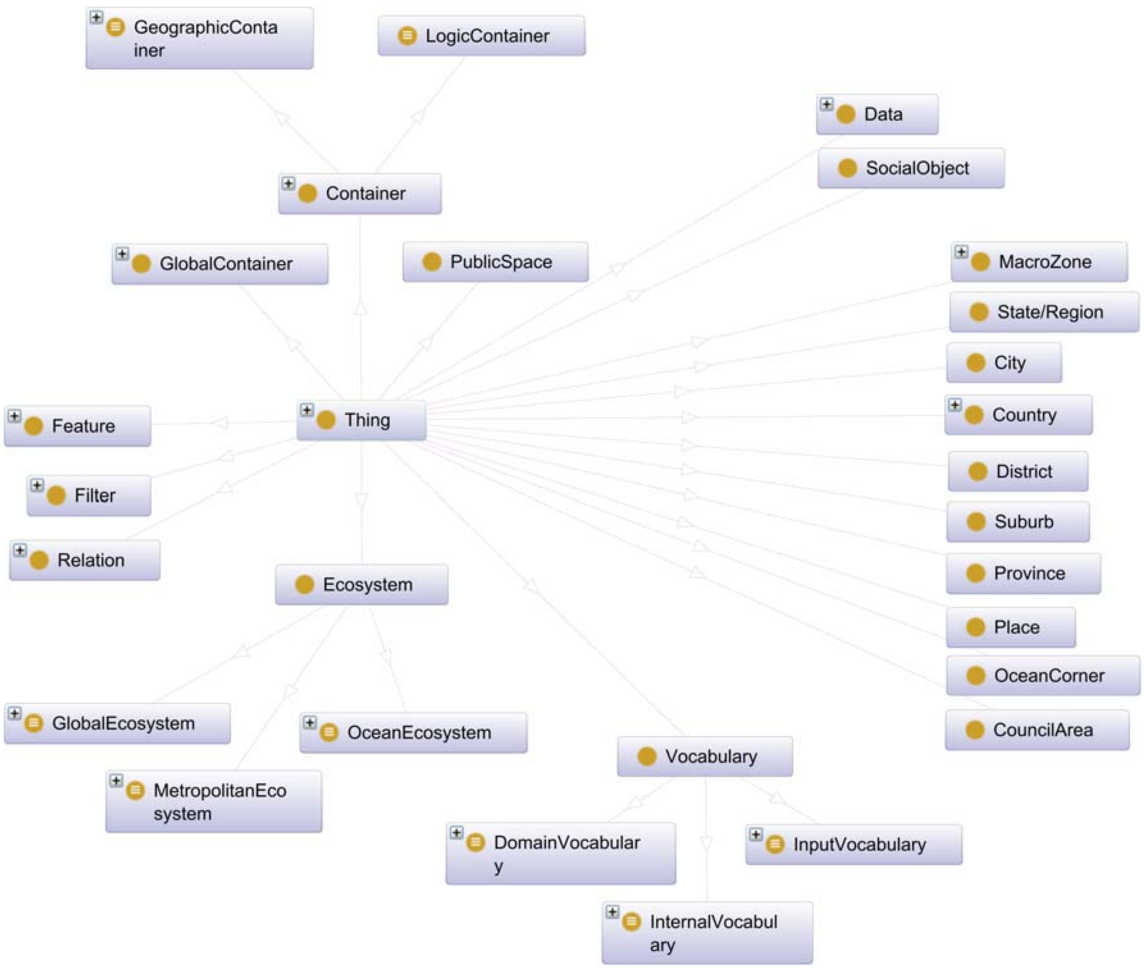

Figure 2: MANSION-GS concept view (core elements).

5. An application Scenario: public Spaces inside Metropolitan Ecosystems

Different ecosystems (e.g. *Metropolitan* and *Ocean Ecosystems*) result from composing different physical containers. The use case proposed in this section focuses on the specification of complex public spaces inside Metropolitan Ecosystems. A Metropolitan Ecosystem mostly describes a City (or parts of it). A public space is a social space that is generally open and accessible to people. Roads, public squares, university campuses, parks, lakes and beaches are commonly considered publics pace. They are very popular and object of interest in the context of several domains and disciplines (e.g. urban design, art, planning and social studies).

In this context, the ecosystem is limited to four physical containers (*Place*, *District*, *MacroZone* and *City*) logically embedded according to a bottom-up focus.

The whole ecosystem is represented by an instance of the MANSION-GS Ontology and is the input for a semantic reasoner (developed over PELLET (Sirin et al. 2007)). The reasoner aims at

the conversion of the knowledge inside the ontology to a human intelligible model (Google Maps, in this case).
Our example refers to the city of Auckland (New Zealand). The city is composed of five MacroZones (Central, North, South, East and West) that mostly reflect the official view of the city for authorities and citizens. As common, the city also can be divided in Districts (or Suburbs), each of them associated to one or more MacroZone. In the same way, Places are composing Districs/Suburbs and, indirectly, the MacroZones and the City. *Public Spaces* (Figure 4, up) are composed of containers (normally *Places* or *Part of Places*) and of atomic elements (e.g. *Buildings*).
Apart from the data, the input vocabulary is composed of a set of relations and features. It is represented in the Table I. Features are used in order to provide a semantic classification of the spaces. The relations from the input vocabulary integrate the generic set provided by the model.
The whole understanding of the space is provided by a multidimensional model in which each relation defines a single dimension. An example is shown in the Figure 3, where the diagram represents the relations among the different spaces composing a subset of the target space according to the model. It is represented in Figure 4 (down).

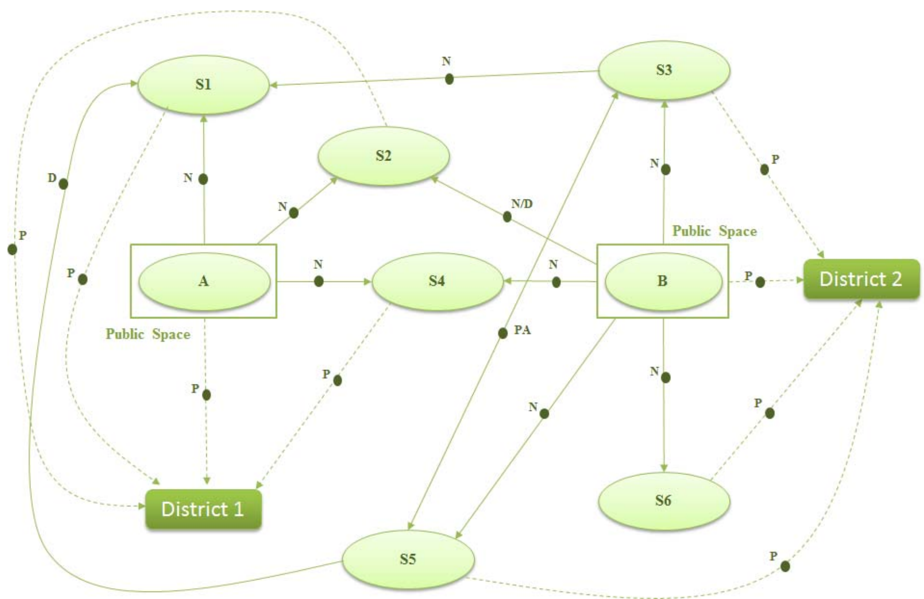

Figure 3: Diagram of the relations for a subset of the target space.

In practice, each element is classified according to the input domain and containers are related to each other by using semantic relations and features.
Domain specific data is imported into the system and semantically related to containers as well as to a basic flow of data from different social media. Data is organized according to several data layers (e.g. tourism).

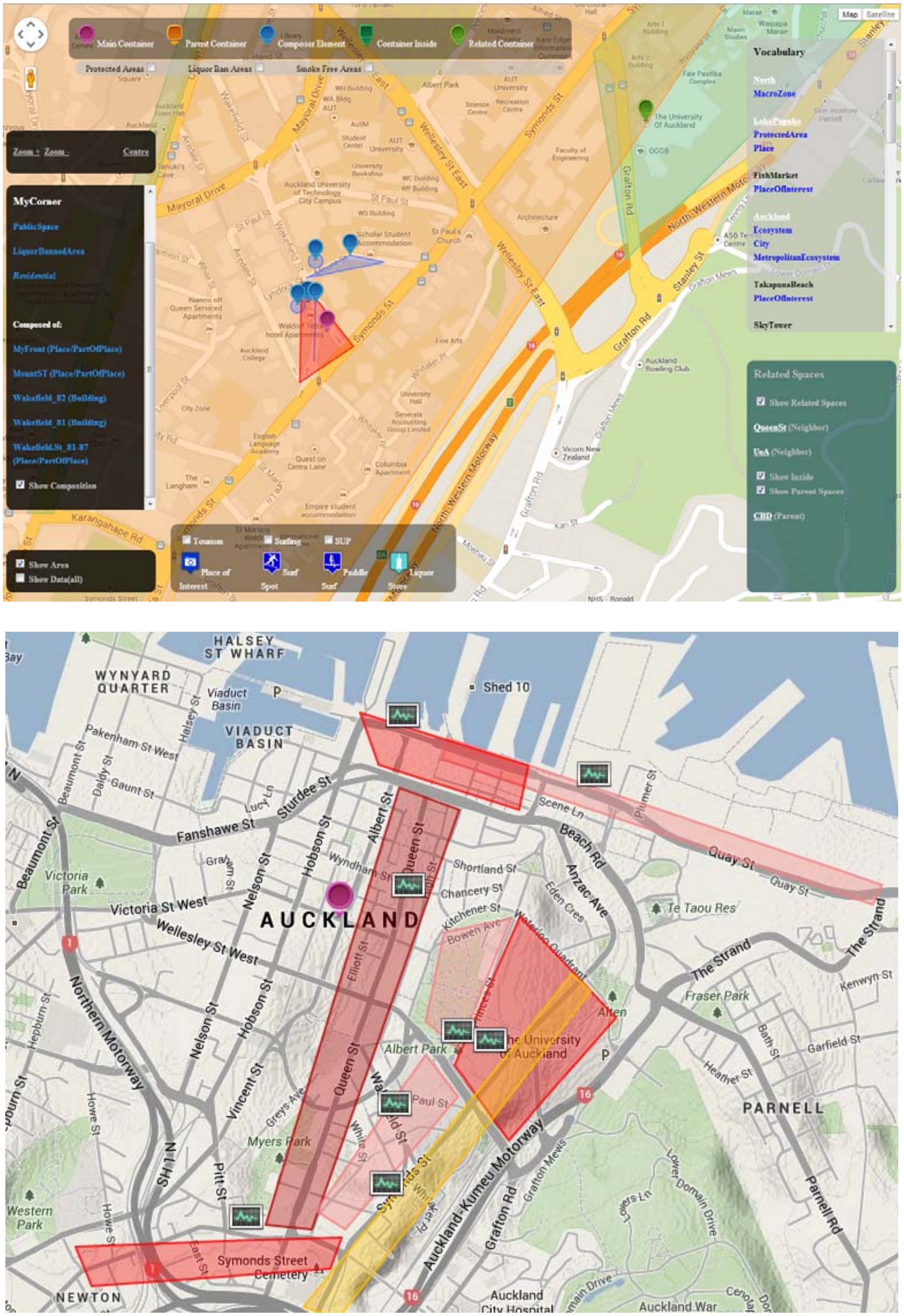

Figure 4: Example of dynamic public space inside the city of Auckland (up) and view of a subset of the overall space (down).

Input Vocabulary			
Concept			Description
Concept	Relation	***Parent (P)***	From the space model.
		Composing (C)	From the space model.
		Including (I)	From the space model.
		Neighbor (N)	It refers to a physically continuous (or close) space.
		Dependent (D)	It defines dependencies among spaces.
		Pair (PA)	It marks a semantic similitude among spaces in function of some parameter/feature.
	Feature	***Residential***	Properties associated to containers that are integrated with specific semantic profiles at application-level.
		Metropolitan Area	
		Commercial	
		Beach	
		Entertainment	
		Lake	
		Nightlife	
		Mountain	
		Transit	
		Natural Reserve	
		Education	

Table 1: Input Vocabulary.

6. Conclusions

Semantic models are a direct and effective way to integrate common environments with semantic capabilities. Semantics act as the n-th dimension on the space providing the ideal convergence point for multiple perspectives of the same reality as well as open relations among composing elements.

Furthermore, data and social objects are embedded in the same data model representing a dynamic ecosystem in which complex dynamics and trends can be defined. The proposed approach is domain and application independent that provides a logic framework in which semantic perspectives of the space can be defined according to applications' requirements. This last aspect could be extremely relevant in a multi-application environment in which the coexistence (and eventual cooperation) of different services and applications could be developed and work in a semantic-interoperable context.

References

Pileggi, S. F., Calvo-Gallego, J., & Amor, R. (2013, September). Bringing Semantic Resources Together in the Cloud: From Theory to Application. In*Computational Intelligence, Modelling and Simulation (CIMSim), 2013 Fifth International Conference on* (pp. 113-118). IEEE.

NG, T. A., Vinh, P. T., & Duy, H. K. (2012, August). A Study on 4D GIS Spatio-Temporal Data Model. In *Knowledge and Systems Engineering (KSE), 2012 Fourth International Conference on* (pp. 34-38). IEEE.

Mao, X., & Li, Q. (2011, June). Ontology-based web spatial decision support system. In *Geoinformatics, 2011 19th International Conference on* (pp. 1-4). IEEE.

Strintzis, M.G.; Mademlis, A.; Kostopoulos, K.; Moustakas, K.; Tzovaras, D., "A Novel 2D Urban Map Search Framework Based on Attributed Graph Matching," MultiMedia, IEEE , vol.PP, no.99, pp.1,1, 0. 2009.

Zhang, C., Zhang, X., Jiang, W., Shen, Q., & Zhang, S. (2009, December). Rule-based extraction of spatial relations in natural language Text. In*Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on* (pp. 1-4). IEEE.

Ardizzone, E., Di Miceli, F., La Cascia, M., & Mazzola, G. (2012, November). Extracting Touristic Information from Online Image Collections. In *Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on* (pp. 482-488). IEEE.

Luckel, F., & Woloszyn, P. (2009, July). A «perlaborative» environment for sustainable cities design staff in a participative perspective. GIS and knowledge database. In *Computers & Industrial Engineering, 2009. CIE 2009. International Conference on* (pp. 1700-1705). IEEE.

Shahabi, C., Banaei-Kashani, F., Khoshgozaran, A., Nocera, L., & Xing, S. (2010). GeoDec: A framework to visualize and query geospatial data for decision-making. *Multimedia, IEEE*, *17*(3), 14-23.

Eldien, H. H. (2009, July). Noise mapping in urban environments: Application at Suez city center. In *Computers & Industrial Engineering, 2009. CIE 2009. International Conference on* (pp. 1722-1727). IEEE.

Wiegand, N., Zhou, N., & Cruz, I. F. (2003, July). A web query system for heterogeneous geospatial data. In *Scientific and Statistical Database Management, 2003. 15th International Conference on* (pp. 262-265). IEEE.

Gao, H., Zhang, H., Hu, D., Tian, R., & Guo, D. (2010, June). Multi-scale features of urban planning spatial data. In *Geoinformatics, 2010 18th International Conference on* (pp. 1-7). IEEE.

Gomes, J., Urbano, P., Montenegro, N., & Duarte, J. (2012, June). A computer-aided urban planning tool driven by semantic web ontologies. In *Information Systems and Technologies (CISTI), 2012 7th Iberian Conference on* (pp. 1-6). IEEE.

Zheng-yu, D., & Quan, W. (2009, April). Road network analysis and evaluation of Huizhou City based on space syntax. In *Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on* (Vol. 3, pp. 579-582). IEEE.

Zhang, Y., Zhao, H., & Li, H. (2010, October). The study on city general planning and management information system based on GIS. In *Computer Application and System Modeling (ICCASM), 2010 International Conference on*(Vol. 8, pp. V8-131). IEEE.

Muramoto, T., Nakayama, K., Kobayashi, Y., & Maekawa, M. (2006, October). Resident Participating GIS-Based Tsunami Disaster Control Systems for Local Communities. In *Communications and Information Technologies, 2006. ISCIT'06. International Symposium on* (pp. 701-706). IEEE.

Suhong, Z., Lijun, Y., & Lifang, D. (2010, August). The Spatial-Temporal Pattern of People's Daily Activities and Transportation Demand Analysis-A Case Study of Guangzhou, China. In *Management and Service Science (MASS), 2010 International Conference on* (pp. 1-4). IEEE.

Yue, Z., Fengri, L., & Weiwei, J. (2011, September). Analysis on land use change and landscape pattern in Hailun city. In *Electronics, Communications and Control (ICECC), 2011 International Conference on* (pp. 1886-1889). IEEE.

Pileggi, S. F., Fernandez-Llatas, C., & Traver, V. (2012). When the Social Meets the Semantic: Social Semantic Web or Web 2.5. *Future Internet*, *4*(3), 852-864.

OWL-Web Ontology Language (OWL), http://www.w3.org/2001/sw/wiki/OWL.

Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., & Katz, Y. (2007). Pellet: A practical owl-dl reasoner. *Web Semantics: science, services and agents on the World Wide Web*, *5*(2), 51-53.

Google Maps API, https://developers.google.com/maps/.

Pileggi, S. F., & Amor, R. (2013). Addressing Semantic Geographic Information Systems. *Future Internet*, *5*(4), 585-590.

Janowicz, K., Scheider, S., & Adams, B (2013). A Geo-Semantics Flyby. In Proceedings of the 9th international conference on Reasoning Web: semantic technologies for intelligent data access (RW'13).

Koubarakis, M., Karpathiotakis, M., Kyzirakos, K., Nikolaou, C., & Sioutis, M. (2012). Data Models and Query Languages for Linked Geospatial Data. In*Reasoning Web. Semantic Technologies for Advanced Query Answering* (pp. 290-328). Springer Berlin Heidelberg.