An Ontology for Ecological Urbanism: SUM+Ecology

Berta Cormenzana, Ferran Fabregas, Maria-Cristina Marinescu, Mónica Marrero, Salvador Rueda, Rosario Uceda-Sosa

{bertacormenzana, ferranfabregas, rueda} @bcnecologia.net, {mariacristina.marinescu, monica.marrero} @bsc.es, rosariou@us.ibm.com

Abstract
As the complexity and abundance of city data increases, reusable semantic models that can integrate heterogeneous data sources in a lightweight manner enable a holistic view of the city data, which is key to Urban Ecology. Our multidisciplinary team has built an ontology for Urban Ecology that not only captures a field-validated urban model and certification process, but also explores the reuse of semantic models and their interaction with domain experts.

Introduction
According to the World Health Organization, by 2010, well over 50% of the World population lived in cities. This proportion will increase to 60% by 2030 and to 70% by 2050 [WHO]. The unquestionable process of urbanization has resulted in an explosive occupation of the land and an isolation of natural areas, with the consequent loss of biodiversity, distortion of the water cycles, galloping consumption of resources and amount of pollutant emission. Our best way to resolve these dysfunctions is to leverage information technology and create a semantic city model for urban sustainability [Town] [Gold].

The innovative knowledge model of a sustainable city [Rueda, 1995] by the Urban Ecology Agency of Barcelona (BCNEcologia-M) is based on two principles: (a) the efficiency of urban systems and (b) the interrelated ecosystem that defines urban habitability. The first principle predicates that the permanence of urban ecosystems depends on the tendency to a higher level of organizational complexity while resource consumption is reduced, just like any other complex system (e.g., the human body, animal and human societies).

In urban systems, the guiding function to measure efficiency (Rueda, 1995) can be expressed as \( \left( \frac{E}{nH} \right) \) (Fig1), where \( E \) is the consumption of energy (as a synthesis of the consumption of resources), \( n \) is the number of urban legal entities (businesses, institutions, infrastructures, and associations) and \( H \) is the value of the diversity of these legal entities which is also known as urban complexity (organized information). With the tendency in today’s societies to produce bigger cities, the values of the guiding function get higher over time instead of lower, as it should be.

The Ecological Urbanism Model
Ecological Urbanism is based on the idea that the permanence of urban ecosystems is guaranteed if there is a tendency to a higher level of organizational complexity while resource consumption is reduced, just like any other complex system (e.g., the human body, animal and human societies).

According to the World Health Organization, by 2010, well over 50% of the World population lived in cities. This proportion will increase to 60% by 2030 and to 70% by 2050 [WHO]. The unquestionable process of urbanization has resulted in an explosive occupation of the land and an isolation of natural areas, with the consequent loss of biodiversity, distortion of the water cycles, galloping consumption of resources and amount of pollutant emission. Our best way to resolve these dysfunctions is to leverage information technology and create a semantic city model for urban sustainability [Town] [Gold].

The innovative knowledge model of a sustainable city [Rueda, 1995] by the Urban Ecology Agency of Barcelona (BCNEcologia-M) is based on two principles: (a) the efficiency of urban systems and (b) the interrelated ecosystem that defines urban habitability. The first principle predicates that the permanence of urban ecosystems depends on the tendency to a higher level of organizational complexity while resource consumption is reduced, just like any other complex system (e.g., the human body, animal and human societies).

In urban systems, the guiding function to measure efficiency (Rueda, 1995) can be expressed as \( \left( \frac{E}{nH} \right) \) (Fig1), where \( E \) is the consumption of energy (as a synthesis of the consumption of resources), \( n \) is the number of urban legal entities (businesses, institutions, infrastructures, and associations) and \( H \) is the value of the diversity of these legal entities which is also known as urban complexity (organized information). With the tendency in today’s societies to produce bigger cities, the values of the guiding function get higher over time instead of lower, as it should be.

Copyright © 2014, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
The BCNEcologia Model applies a systemic approach to the management of cities to encourage a more sustainable model and create a universal tool, applicable to any urban system, as a protocol for measuring the degree of efficiency and quality of cities (Fig 3).

This innovative certification tool [RCV], unlike other existing certifications as LEED for Neighborhood Development [LEED] or BREEAM Communities [BREEAM], addresses all aspects of the urban ecosystem, not just the physical environment: (1) it is a system that is based on an intentional model, (2) it is developed under a systemic view of ecosystems and (3) it is applicable to different scales (city, district, urban fabric) and urban processes (new urban development and urban fabrics in use).

The methodology for the assessment and accreditation of sustainability is specified through a system of indicators and it is materialized with a seal of urban quality, guaranteed by the Spanish Government Building Ministry.

The BCNEcologia Model establishes a common framework for the communication between the city council, city planners, official and private institutions, academic centers, and society. It constitutes a reference tool to aid in the urban planning process and to guide the configuration of smart cities.

BCNEcologia-M has been successfully applied to both neighborhoods and cities, to new construction projects as well as to the study and measurement of the degree of sustainability in Spanish cities of various sizes, including Barcelona, Vitoria-Gasteiz or Sevilla. Currently the model and certification process is being evaluated by Moscow (Russia) and Quito (Ecuador).

The Ecological Urbanism tool to certificate the degree of fitness of any analyzed proposal to the most sustainable city model is parameterized by a package of 50 indicators [RCE]. The model and its associated process and indicators are organized in seven dimensions:

- **Land Occupation.** E.g., Housing density, Absolute compactness.
- **Public Space and Habitability.** E.g., Air Quality, Thermal comfort, Habitability in public space.
- **Mobility and Services.** E.g., Proximity to alternative transportation networks, Parking space for bicycles, Work self-containment.
- **Urban Complexity.** E.g., Urban diversity, Knowledge intensive activities, Spatial and functional street continuity.
- **Green areas and Biodiversity.** E.g., Soil biotic index, Proximity to green spaces.
- **Urban Metabolism.** E.g., Energy self-sufficiency from renewable energies, Greenhouse gases emissions, Water sufficiency, Closing the organic matter cycle.
- **Social Cohesion.** E.g., Synthetic index of social inequality, Proximity to public facilities, Citizen Participation in urban processes.

As it becomes apparent, the paradigm of Ecological Urbanism integrates the physical infrastructure (e.g., buildings, mobility networks), the people, the legal entities and their activities, together with the information flows. Up until now, models and methodologies that capture the urban reality have addressed only vertical, narrow-focused issues and have missed on the opportunity to integrate them transversally.

The principles of Ecological Urbanism are directly linked to the Entity Model of the city (Figure 2) but they reflect a need to establish a reusable City ontology with a generic vocabulary and semantic context to (a) facilitate the interoperability of sub-models and urban classes, (b) manage the evolution of the model itself and (c) manage the model extensions when implemented in different cities with different and heterogeneous data sources.

**The BCNEcologia Certification Process**

The computation of the sustainability indicators of BCNEcologia-M is currently composed of three phases: (1) Data processing and integration into the Geographic Information System (GIS) and simulators, (2) Calculation process that includes simulation, spatial analysis (proximity, overlay, statistics, and network analysis among others) and map layout generation, and (3) The Evaluation Process which involves filling predefined forms to extract the final values.

A City ontology that captures both the underlying entities and the certification indicators can act as a ‘smart
whiteboard’ where entities, their data, and their relations are kept globally, helping with the validation, navigation, and exploration of the data by domain experts (Fig3).

Fig3. BCNEcologia Certification Template.

The SUM+Ecology Ontology

In domains where a complex, interrelated system is partially captured by heterogeneous data sources, like in the case of the human body [SCH] or a city, ontologies can act as a lightweight integration semantic layer that indexes and categorizes these data sources to provide a holistic view of the system without altering them. General semantic models can also help in standardizing and reusing vocabulary and relations among domain entities [KMc].

To help realize these goals, we have implemented the SUM+Ecology ontology as an extension of SCRIBE [SCR], an ontology for Smart Cities which already describes events, stakeholders (people, institutions), physical infrastructure (buildings, streets, sensor networks, etc.), and indicators (pollution, traffic flow, budget, etc.) as shown in Fig 4. We have populated and validated SUM+Ecology with data from the City of Barcelona and we are currently developing a tool for its navigation (http://redrock.bsc.es:8080/SemanticUrbanModel2/master.jsp) and query to be discussed in the next section.

With this work, we expect to speed up and standardize the BCNEcologia certification process, as opposed to the current semi-automatic population and manual extension of the model. We also want SUM+Ecology to provide integrity to city data by using the semantic patterns below.

- **Data Consolidation.** E.g., two data sources, Finques and Parcelas describe both the taxable features (number of residents) and the physical infrastructure (area) of land lots. This means that data properties like TotalBuiltArea may be semantically associated to more than one concept, e.g., BuiltStructure, Building or Landlot. Our ontology allows associating TotalBuiltArea to say, LandLot while allowing access to it through related entities. As a result we can consolidate and even validate data sources without altering them.

- **Semantic Categorization of Enumerated Types.** E.g., BuildUseUnits (a dwelling within a building) are assigned tags that describe its business use (commercial, residential, a workshop, etc.) or whether it’s leased. There are over 50 tags related to business uses in BCNEcología-M. Mapping them to a semantic graph with relations and abstract categories which has been obtained from the International Standard Industrial Classification of All Economic Activities [UNIC], enables queries with a richer vocabulary.

- **Reuse-and-Refine** of data properties. The data property numberOfResidents is usually associated to a building, but in the real city data it’s a field of LandLot because this is how the city keeps track of this information due to tax purposes. By making numberOfResidentsPerLandlot inherit from numberOfResidents, we can maintain semantic integrity and still query for numberOfResidents associated to a building through reasoning, even though the underlying model does not directly contain the information.

These patterns are part of the SCRIBE set of design principles. Other principles include (1) using a simple dialect of OWL (OWL-QL) [OWL2] without constraints, (2) categorizing object properties along UML-Like relations, so inferences can be performed at an abstract level without knowing the details of the extensions and (3) heavily annotated classes and properties with explicit mapping information to standards or other models.

All these principles are relevant to how the BCNEcologia-M mapping is leveraged. In particular, mapping data from other models, be it a data source (RDB, XML, JSON, RDF/RDFS/OWL), or a standard (NIEM, MISA/MRM, CAP), is relatively easy to conceptualize because there are three ways in which SCRIBE stores simple data types. (1) Data from simple types (strings, Booleans, integers, etc.) is stored in data properties. E.g. NumberOfFloors = 15 for Building Albatross. (2) Data from enumerated types is represented by classes or hierarchies, which we call Enumerated Type Classes, or (ET-) Classes. E.g., AssetUse (where asset may be a BuildingUseUnit, an atomic dwelling in a building) has children ‘CommercialUse’, ‘StoreUse’, etc. These classes are associated to singleton instances, like ‘CommercialUse_INST’ and ‘StoreUse_INST’, which are part of the model and are called Value-Type (VT-) Instances. An instance, say, the individual dwelling ‘Albatross BajoA’ is related to one or more of these VT-Instances through an object property defined in the model, as shown in Fig 5. VT-Classes organize enum types in hierarchies, which is particularly useful in large types like the use of an asset (commercial, industrial, government,
etc.). The SCRIBE principles require that the only relation between concepts and instances is \texttt{rdf-type}. We introduce VT-Instances to ensure the delimitation between the class and instance graphs for the purpose of inferencing efficiency. Finally, (3) Relations to other (non VT-) instances.

**Fig 4. The SCRIBE Model for Smart Cities**

The atomic data in the model can be grouped as one of three types of entities, i.e., pairs of type 
\(<\text{name}, \text{atomicValue}>\), \(<\text{objectProperty}, \text{VT-Instance}>\), and \(<\text{objectProperty}, \text{Instance}>\). Each of these pairs, \(p_i\), can be accessed directly through a set of classes (entities) of the model \(M\), which constitutes its semantic context, \(C_M[p_i]\). We can comparatively study the organizational complexity and richness of two models by analyzing the mapping of their atomic data contexts. In the next section we analyze the mapping from BCNEcologia-M to SCRIBE.

**Analyzing Semantic Model Mapping.** The SUM+Ecology ontology is an extension of the SCRIBE core ontology \([\text{SCR}]\) with an independent namespace for classes (\(\text{BCNEcology}::<\text{}\)), properties, and VT-Instances. The process of mapping the current model of BCNEcologia has been systematically carried out by mapping the three atomic data pairs described above, and then mapping the entities to classes. (1) \(<\text{name}, \text{atomicValue}>\) pairs are mapped to data properties in SCRIBE, or added if they don’t exist in the core model. (2) Enumerated types are mapped to class hierarchies and their VT-Instances created if needed, and (3) classes are semi-automatically mapped by resolving \(<\text{name}, \text{atomicValue}>\) and \(<\text{objectProperty}, \text{VT-Instance}>\) pairs. Following the SCRIBE principles, we have kept provenance and mapping information as metadata in classes and object properties. Examples of the mapping are shown in the next section.

The current SCRIBE model has 1407 classes and property types, and 1799 VT-instances. The BCNEcologia mapping of city entities includes 90 classes and property types, and only 8 were added in the \(\text{BCNEcology}::<\text{}\) namespace. The 50 indicators of BCNEcologia-M were also added, as SCRIBE did not include specific urban indicators. We have also mapped 91 instances that represent values of enumerated types. Of these, 65 were already in the SCRIBE Core.

**Fig 5. City Entities High Level Model.**

We have computed the context of each BCNEcologia-M mapped pair \(C_{\text{sc}}[p_i]\), where \(\text{SC}\) is the SCRIBE deductive closure with respect to the OWL Micro reasoner in Jena. The size of these contexts, after eliminating the topmost concepts of the SCRIBE ontology, which are used for general organization of the ontology, is, on the average, 4.228 larger than in BCNEcologia-M with an average hierarchy depth of 3. This means that SUM+Ontology is considerably more dense in relations than BCNEcologia-M, so classes can be accessed through many more navigation patterns and queries. A fragment of the computed context mapping is shown in Fig 6.

**Fig 6. BCNEcologia-M properties (blue) mapped to SCRIBE concepts in their semantic context (red).**

It is encouraging that the classes in BCN-Ecologia could be easily embedded in SCRIBE, given their similar domains and the richness of the SCRIBE vocabulary. Each of the 33 (non ET-) classes directly mapped from BCN-Ecology can be accessed by a path of 3 or less an average of 6.2 different ways, from each other, when considering all of the SCRIBE model—not just the BCNEcologia-M mapped classes. This means that SCRIBE model considerably enrich the original vocabulary.

These purely topological measures need to be complemented with usability studies. I.e., the complexity of the resulting model needs to be justified in terms of its efficacy in helping domain experts query the information. In the sections below we discuss our current prototype and our plans to conduct these studies in the next few months.
SUM+Ecology Visualization and Query

We’re currently developing an ontology tool with two goals: (1) To integrate information (including geopositioned data) without changing it, through a model that can be easily extended and reused; (2) to offer an ontology visualization and exploration tool for domain experts that doesn’t require understanding the underlying (ontology and query) technologies. This section focuses on the last goal.

Our current work aims to bridge the gap between the ease of use of web search tools and the more sophisticated patterns in ontology querying by helping users to (a) build simple queries in an intuitive way, (b) build complex queries from simple ones, and (c) understand the content of the ontology. To address the first two challenges, our basic goal is that the expert user shouldn’t have to learn SPARQL to be able to query the data. As it will become apparent in this section, the simple exploration scenarios can be seen, at a high level, as a Keywords + Constraints query, as in a Web search engine. To construct complex queries we adopt a dialog-based, iterative approach in which as many queries and constraints may be defined and updated as needed.

Challenge (c) takes the user centric view a step further to assist those domain experts which, besides being query language agnostic, are neither willing to work with the ontology nor to understand its details. For these users it’s important to easily explore the ontology for concepts and attributes to be queried. We provide this functionality via two mechanisms: keyword-based search for properties (and the classes they pertain to), and contextual navigation of the model to explore sets of related classes.

One of the guiding principles of SUM+Ecology is to discourage the definition of constraints within the model and use OWL-QL for inference. Validation must therefore be done programatically by the applications, and the querying functionality will work on the data as specified without considering the constraints that may be defined as part of the model. This strategy ensures that querying is decidable, fast, and flexible. In what follows, we describe the tabs and its capabilities for domain expert queries.

Enabling ontology queries by domain experts. The tool we are currently developing uses exclusively open source technologies, which shows semantic web technologies are reaching maturity and can be used to handle real city data. We’ve used OpenLayers v.2.13.1 for the map functionality, OpenStreetMap for the maps themselves, Apache Tomcat 7.0.47 for the web server, Apache Jena v.2.11.0 library for ontology management, and Parliament v.2.7.4 for managing and querying geospatial instances in the ontology. Parliament is a data management solution that is compatible not only with OWL and SPARQL, but also with GeoSPARQL. It implements temporal and geospatial indexing and includes an RDFS/OWL simple inferencing engine. We’ve used City of Barcelona data to test and validate the tools. We visualize domain entities as linked semantically, not just by their geospatial colocation, as in the today’s BCNEcologia-M.

![SUM+Ecology Visualization Tool](image)

**Figure 7. SUM+Ecology Visualization Tool**

Figure 7 shows a fragment of the SUM+Ecology tool opened at the **Data-Query** tab, designed to help domain experts iteratively filter a query, as discussed below. There are other three tabs: **Urban Concepts**, allowing class search and navigation, Related-Concepts, enabling a generic navigation of object properties along UML relations (Fig 8) and Sparql-Query, which allows free format SPARQL queries. The map manipulation is done through large-grained geospatial entities in the ontology: Blocks or Landlots. An arbitrary polygon may be selected, and instances will be filtered through the selection of classes in the **Urban Concepts** tab. User-selected instances appear in the **Data-Query** tab, as a context for domain expert queries, along with relevant model information.

The **Class Window** displays the classes that have instances selected on the map. This is the result set S. The instances and their properties are visualized in the instance window. Users can also visualize data available in a type of instance at all times, either in the **<Property>** button (Fig.7) or in Related Classes in Fig 8. From this point on, we think of the interaction of the domain expert as a dialog that further refines S through several actions:

- **Query** for the result of a collective operator applied to a data property of either one of the classes c in S or a class c’ that is related to c via an object property. Let Q be the instance set of c on which we apply the **collective operator**. The collective operations we currently support are average, sum, min, max, and count. E.g., the **oldest building in an area**, or the **average building height**.
- **Filter** the set of instances in Q using conjunction (AND) and disjunction (OR) operators. Conjunctive operators apply sequentially to Q to obtain a subset of instances Q1. With each definition of a disjunctive operator a new subset of filtered instances (Q2,
Q3…Qn) is created starting from Q. The final result set R is Union(Q1, Q2,…Qn). Supported value constraints are comparisons with the value of the properties of c or c’. An example of constraint is, for instance, filtering for those public buildings that provide health services. The GUI helps visualize the effect of enlarging or restricting the set of constraints by the simple action of checking and unchecking a set of buttons associated with the different conditions (Fig 7). E.g., the buildings in an area which have been built after 1970 and are higher than 20 meters.

- Define any number of the previous queries, each one over a potentially different class c in S, and compute aggregate results by applying arithmetic operations over them. E.g., average living surface per inhabitant of the buildings in the area which have been built after 1970 and are higher than 20 meters.

- While this approach is less powerful than direct SPARQL querying, being able to define, update, and reset filters, property names, and types of analytic operations in a dialog style is an intuitive way for domain experts to query the ontology.

The Related-Concepts tab has a double functionality. We can search for properties as keywords to navigate the context of classes in the ontology, so one or more entities can be selected, filtered, and jointly queried. We also provide contextual navigation to include sets of related classes using high-level, abstract properties (associated-to, has-aggregate, aggregate-of, has-attribute, attribute-of). This is an iterative process that which starts with the selection of urban concepts within a geographical area and navigates through their properties to concepts of interest. The functionality enables the tool to find information related to a class even when the attribute required is associated with another class related to it. E.g., in our scenario, TotalBuiltArea may be associated to a LandLot, not a Building, depending on how a city stores the data. The TotalBuiltArea of the LandLot where the building is (associated-to) is shown as the semantically closest entity.

Related Work. At the expressive end of the spectrum, the OWL-DL tool for ontology browsing and querying, OntoIQ (Ontoligent Interactive Query Tool) [OntoIQ] is based on composable query patterns concept, role, and, union, and combo. The combo pattern enables the construction of nested queries based on simpler ones, and users must provide additional ontology-specific information and confirm the translation into the RACER Query Language (nRQL)[nRQL] before they are evaluated by the reasoner. A challenge with languages based on predefined query patterns is that they don’t exhaustively cover all useful patterns. Also, patterns impose a strict discipline for defining queries, which may turn out not to be intuitive. Systems such as those developed at the University of Manchester [Man1, Man2], the KSL laboratory at Stanford [KSL], or University of Maryland (not publicly available) adopt approaches that support (some, partially) the DQL specification, use powerful reasoners, but don’t facilitate domain expert queries.

Fig 8. Abstract navigation of relations

At the other end of the spectrum, visual querying focuses on usability at the expense of expressivity and raises scalability issues in terms of the complexity of formulating the queries to capture the needed information. Faceted search (FS) is a form-based search approach based on a series of orthogonal dimensions to be combined to filter the solution space and are used by online sellers like Amazon. Query by navigation (QbN) allows users to navigate relationships between objects in a knowledge graph, and is used in Web browsing. Both these approaches are more appropriate for instance navigation than computing analytic results over instances, like in the Urban Ecology domain. Optique[Opt] is a European project that takes a visual, ontology-based approach for scalable querying, based on mesh-ups of widgets, one for concept navigation via relationships, another for defining constraints over the attributes of the currently active concept. These may be combined depending on the query context and experience of the user.

Although not ontology-based, web search tools are successfully –useably- solving key issues for not IT-users: How to easily formulate the query, how to improve precision and recall of query results, and how to support efficient, exploratory queries in an open world.

Conclusions and Future Work

The goal of the SUM+Ecology project is to create an ontology, supporting tool, and a set of map and extension patterns to facilitate the Urban Ecologism assessment of medium and large sized cities by domain experts by January 2016, based on a validated, real life model.

We started in November 2013 and we already have a prototype to test the usability of our query and visualization strategy in the next few months. Our next phase is the creation of a semi-automatic pipeline that allows domain experts to map data sources from new cities to the ontology. Also, efficient population of high volumes of data and usability studies will be the focus of our work into 2015.
References


[nRQL] V. Haarslev, R. Moeller and M. Wessel (2004) Querying the semantic web with RACER + nRQL. CEUR Workshop on applications of description logics (ADL04)


