GrowSmarter
Transforming cities for a smart, sustainable Europe

FACTSHEET

Big Open Data Platform

PART OF SMART SOLUTION 8: BIG DATA MANAGEMENT

INTEGRATED
INFRASTRUCTURES

- Integrates data from a variety of domains to form one single access point for application developers.
- Provides a universal and accessible platform that allows application developers to build uniform services for cities with minimal effort and no redesign.
- Promotes equal access to public data, fostering sustainable and equitable public development

Barcelona

- Technical partner: Barcelona Supercomputing Centre
  Contact: maria.marinescu@bsc.es

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement no 646456. The sole responsibility for the content of this document lies with the author and in no way reflects the views of the European Union.

GrowSmarter
Transforming cities for a smart, sustainable Europe

What is the solution?

Our solution consists of developing a semantic model that reflects and connects three domains of interest: mobility, energy, and integrated infrastructures.

Users can browse and query the ontology. A semantic access layer will translate semantic application queries (i.e. queries over the semantic model rather than the actual data) to queries that accesses the city data platform.

One of our goals is to provide a solution in which application writers could deploy their services in different cities without modifying the code (if the new city adopts our city semantic model and develops their own semantic access layer).

An important part of this translation involves a cooperative and semi-automatic mapping tool which will return recommendations for correspondences between the semantic concepts and the actual concepts in the city schema.

The goal is to provide a solution that is easier to evolve, maintain, and port to new cities with different data and use patterns.

This should also be a solution that allows applications to work as-is over new data rather than having to undergo implementation changes – to the extent that data exists.

Our approach is based on developing an urban semantic model, concretely an ontology (8.2). An important part in the process of adopting a semantic solution is how difficult it is to populate the model with actual data.

Measure 8.3 semi-automates the data mapping process to our urban semantic model for data monitored in the city of Barcelona.

Measure 8.2 also provides data exploration and semantic access capabilities to the actual data integrated by the GrowSmarter Platform and accessible via a REST-type interface (REST API).

How does it work?

This solution consists of three components:

City ontology, together with a browse and query tool:

The city ontology reflects the meaning (i.e. semantics) of all the urban concepts (entities and relationships) that describe the domains of interest and the connections between them. The browse and query tool allows keyword-based search of concepts, navigation starting from these anchor concepts, and the construction of queries in a graphical fashion.

Semi-automatic mapping tool:

This tool aligns the semantic model and the specific model of the city data platform, and will be available via web. Multiple users could participate collaboratively to produce valid alignments.

Semantic access layer (SAL):

Functions as an access point for applications that pose semantic queries to access the data on the city platform. Applications accessing data from different cities can work without modification if a

SAL exists for those cities that maps between the city ontology and the actual city schema. SAL acts in behalf of the applications (with their security and privacy credentials defined by Cellnex) to fetch the required data via a REST API and compute the query results for the few most common query operations (such as join). It calls the mapping tool to know which are the resources in the city platform schema that correspond to the semantic concepts contained in a query.

Data integration solutions traditionally imply a data warehouse approach. While this is based on a well established and efficient technology, as well as solid formal foundations, several characteristics of data in urban environments are a misfit for this type of data integration.

Firstly, data and schemas evolve; secondly, data is incomplete and no assumptions should be made about non-existing data; thirdly, there are an increasing number of data sources of heterogeneous nature and formats that need to be integrated in an efficient and, as much as possible, automated way; fourthly, data is usually available for consulting but cannot be moved around and stored at the target.

These are scenarios where semantic technologies excel. There are not only a natural fit for the Open World paradigm, but they evolve gracefully and foster semi-automatic mapping techniques for massive data population and access.

One advantage is that new data can be integrated faster, new semantic relationships can be inferred, and users can query the data without having to learn a query language nor understand the entire data model at a time.

Integration with other smart solutions

Our solution accesses city data via the API offered by Cellnex (GrowSmarter API).

Given that this is a REST API, the semantic queries cannot be translated into an SQL query against the city data platform; instead SAL will implement some of the most common operations required by the queries, such as joins.

The GrowSmarter API implements secure and private access to resources.

Expected Impact

The urban semantic model is a transversal solution that can affect all the other measures proposed in GrowSmarter, and has both economic and social (indirect) impact by enabling the query and analysis of integrated city data (as explained above as benefits).

We are focusing on Barcelona data, and we are discussing possible deployment of at least one of the measures in Köln. The idea is that follower cities could leverage our data integration, access, and query tool to access their city data, if this is recorded in digital form and available on a city platform.

The long-term impact is that application developers that want to build services in these cities can do so with little initial effort and no redesign as long as the city platform API doesn´t change and they use query functionality that is implemented as part of SAL. This solution promotes equal access to non-private data and therefore

fosters sustainable and equitable economic development.

Environmental benefits:

Integration of data from different domains via a model that explicitly formalizes the relationships between concepts enables global and comprehensive analysis. This makes it possible to account for long-range effects between aspects that otherwise may seem unrelated, making for better planning, optimization, and decision making in all respects related to urban environment.

Socio-economic benefits:

- Application writers can develop services using a large range of already unified data.
- The Services using semantic rather than direct data access could work as-is for every city that adopts our semantic city model and develops a semantic access layer to city data.
- Such a framework can result in economic benefits to service developers (and providers), can make data access more egalitarian, and can create jobs. From a social standpoint, the approach can be extended to integrate data extracted from semi-structured or unstructured data, such as text posted by citizens and concerning citizens´ needs.

Potential for replication

Other cities can replicate this solution by:

1. Adopting our semantic model. The browse and query tool is geared toward domain specialists but does not require one to either learn a query language or to understand the whole model at a time.

2. Implementing a semantic access layer that uses our mapping tool to translate from semantic concepts to their actual data. This implies the creation of a data schema for the new city (in .owl format) and the implementation of the desired query features in case that the city platform does not accept queries via – for instance – an SQL access point. A domain specialist should be available to validate the correct mapping recommendations on part of applications.

One potential problem is that new applications may require answering queries over entities that were not modelled, or involving relationships that were not contemplated by the current measures.

This requires that a domain specialist and a semantic technology specialist work together to extend the ontology with the missing concepts. The advantage is that the model is flexible and naturally extensible.

A more serious problem arises in the case that the new types of applications are radically different from the ones that guided the ontology design (although within the same domain of discourse), in which case it may be more efficient to redesign the ontology. The browse and query tool, as well as the mapping tool can remain largely unchanged.