Online Geodatabases as a Source of Data for a Smart City World Model Building

Diamantino Ferreira, Fátima Fazenda
Madeira-ITI
University of Madeira
Campus da Penteada, 9020-105 Funchal, Portugal
diamantinorgf@gmail.com, fatima912@gmail.com

Karolina Baras
Centro Algoritmi
University of Minho
Campus de Azurém, 4800-058 Guimarães, Portugal
karolina.baras@algoritmi.uminho.pt

***Abstract* — Online geodatabases have been growing increasingly as they have become a crucial source of information, with increasing social networks providing location-related features to their users. The aim of this study is twofold. On one hand, as our main contribution, we make an overview of the most significant online geodatabases which aims at understanding what data they contain, how they are managed and updated and who are the major consumers of their data. On the other hand, we shortly describe our previously developed world model and propose a new type of data feeds coming from place-based social networks and online geodatabases. So far, this symbolic world model has been fed with data coming from manual editing and automatic sensor-based updating. We believe that the aggregation of these additional sources of data will provide a more rich description of places of interest in a city and will enable an automatic creation of topological and other types of relations between those places.**

Keywords-online geodatabases; place-based social networks; symbolic world model.

I. INTRODUCTION

The ever more increasing use of social networking applications where the users may add location data to their posts, comments and pictures, created the need for geographic data repositories providing complete and correct data freely, with points of interest, monuments, companies, bars, stores, hotels and restaurants, among others. Currently, there are several online databases providing data about millions of places all over the world. Some of the examples are GeoNames [1], LinkedGeoData [2] and Factual [3]. Social networking applications, such as Facebook, Google+ and Foursquare use their own repositories in addition to the above mentioned; they also provide tools for creating new places, for editing the existing ones [4], adding more details, ranking and so on, all aiming at the construction of the most complete, the most correct, the supreme database of places in the world.

We have been working on a symbolic world model that acts as a repository of data about places people visit or know [5]. This model is built similarly to the human mental model of space in which places are represented by nodes of a graph and there are topological relations among them [6]. Each place is further described by a set of attributes. Each relation also owns a set of attributes if adequate and necessary. Our model not only allows topological relations; it also allows custom types of relations.

In our previous work [7], we showed how this model can be constructed manually by the user through a web application and also how it can be updated automatically by a set of processing modules connected to a WiFi network. Now, we are interested in developing applications that interact with Foursquare and/or directly with one or more online geodatabases to help user to construct and update their symbolic world model. The final step in our project will be the integration of all the three data sources, manual, automatic from sensor networks and semi-automatic from online geodatabases, with our system architecture and further study of the most appropriate implementation for the data repository for which we have identified already a set of challenges.

The remaining of this paper is organized as follows. Section II presents an overview of the most significant online geodatabases and discusses some of their aspects. Section III introduces the symbolic world model and describes two developed modules to act as new sources of data for the world model. Section IV concludes and makes an outlook.

II. ONLINE GEODATABASES

Geography has always been important to the human race. Stone-age hunters anticipated the location of their target, early explorers lived or died by their knowledge of geography, and current societies work and play based on their understanding of who and what belongs where. Applied geography, in the form of maps and spatial information, has served discovery, planning, cooperation, and conflict for at least the past 3000 years of our history [8].

Web mapping systems have gained substantial popularity in the last 15 years, and have been rapidly expanding since the release of OpenStreetMap [9] back in 2004, followed by Google Maps in 2005, Wikimapia in 2006 and Microsoft Silverlight in 2007. Latest releases include improved navigation, Google Fusion Tables, Google Maps Mobile and Wikitude Drive (2010) [10].

It is important to note that map coverage did not only become more widespread in recent years; it has also become significantly more accessible to the average user. Early maps that were restrained to commercial and data providers via satellite images rapidly expanded to the combination of road networks, panoramic street-level imagery such as Google Street View and massive collaborative systems such as

 ISBN: 978-1-61208-363-6

OpenStreetMap. Geographic Information Systems (GIS) represent a jump from paper maps like the computer from the abacus [11]. Within the last thirty years, GIS technology has evolved from single purpose, project-based applications to massive enterprise systems [12]. Such systems are currently being used by businesses, institutions, industry, local governments and the private sector to provide services to clients, manage resources, and to address multiple issues concerning to health and human resources, transportation, public safety, utilities and communications, natural resource, defense and intelligence, retail and many more.

With such a development, new concepts and technologies have arisen over the years. 3D Desktop applications, detailed and interchangeable layers, public data, map makers, map integration with social networks and mobile mapping have also appeared. The existing services have attracted several millions of users, both desktop and mobile. The scope of web mapping applications has widened from purely easy to use consumer-oriented tools to highly specialized applications with GIS functionalities that help solving and optimizing problems in several domains. Despite the advances of web mapping within the last few years, there is still a lot of potential to collaborate, to elaborate, to tell stories with creative new methods and to use the data in useful and interesting new ways [10].

Presently, vast quantities of geospatial data and information at local, regional and global scales are being continuously collected, created, managed and used by academic research, for spatial decision support and location based services. A key aspect to any geospatial solution is to support efficient data maintenance and analysis in a heterogeneous operating environment. This requires highly scalable, highly available and secure database management systems. One of the biggest challenges is integrating time into database representations, another is integrating geospatial data sets from multiple sources (often with varied formats, semantics, precision and coordinate systems) [13].

Current GIS applications, which are optimized to store and query data that represent objects defined in a geometric space, often utilize online geospatial databases as sources of their data. Most spatial databases allow representing simple geometric objects such as points, lines and polygons and can operate with varied data structures, queries, indexes and algorithms. These can support several operations such as spatial measurements (computing line length and distances between geometries), spatial functions, geometry constructors and observer functions (queries which return specific information regarding features such as the location of the center of a circle). They also allow remote querying of results via an Application Programming Interface (API) or even integration with other existing geo databases (such as the LinkedGeoData initiative).

Some examples of online geodatabases include OpenStreetMap (OSM) project [9], the GeoNames project [1], and Factual [3]. The main difference between these is that OSM and GeoNames are both open-source projects, accepting data from thousands of users, while Factual, usually, only maps out missing attributes or adds new ones. OSM features the largest collection of data (over 1 billion nodes), which surpasses other services in this regard.

A. Factual

Factual, launched in October 2009, is an open data platform developed to maximize data accuracy, transparency, and accessibility. It provides access through web service APIs and reusable, customizable web applications. Factual features data sets about local place information, entertainment and information derived from government sources. At the current date, and according to their website, Factual contains data of over 65 million places, which are updated and improved in real-time by Factual's data stack. Factual's API allows for remote access to stored data, through the use of queries. Access to the API server must be requested first with an OAuth authorization standard API key and secret, and has existing frameworks for several programming languages such as C#, Java, PHP, Python and Ruby. Results can also be obtained through HTTP GET requests. According to the factual developer site, it is possible to make different types of queries, based on text or points of interest, direct and reverse geocoding, geo filters, data submission and matching queries with provided information. It can also clean and resolve data as results are submitted into the database, and afterwards connect it to other sources of factual data, or external sites (such as Facebook or Twitter) relating to the same coordinates, and further enhance this by delivering contextually relevant content, experiences and ads based on where mobile users are located.

B. GeoNames

GeoNames is another global geodatabase initiative. According to their website, the GeoNames database contains over ten million geographical names corresponding to over 7.5 million unique features. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes. Beyond names of places in various languages, data stored includes latitude, longitude, elevation, population, administrative subdivision and postal codes. However, implementing its own semantics schema can be a barrier to future interoperability with other geospatial databases and sources of information [14]. It operates on a low level, semantic schema with no set of constraints on the domain and range of attributes, which can be a barrier to the consistency of the data set. GeoNames also features a functional API call server, which allows for querying of information on a programming level in several programming languages (such as C# or PHP). Results can be returned via XML or JSON objects. As described on their website, GeoNames API allows for full-text based searches, postal code search, place name lookup, nearby postal codes, reverse geocoding, nearby populated places as well as providing other services such as recent earthquakes or weather status lookup.

C. OpenStreetMap

OpenStreetMap is a collaborative geospatial project that aims at creating a free editable map of the world created by

Steve Coast in the United Kingdom in 2004. Since then, it has experienced a growth of up to 1 million users (in 2013) who collect all kinds of high quality data from various sources, such as Global Positioning (GPS) devices, aerial photos, sea travels, and government files. Like most Internet projects, most users are casual or inactive, with a small dedicated minority contributing the majority of additions and corrections to the map. The database now contains over 1 billion nodes, 21 million miles of road data and 78 million buildings [15]. OSM features a dynamic map where every urban or natural feature is built by the community, resulting in accurate, high quality data representations [16].

The OSM project facilitates complete, regularly updated copies of its database, which can be exported and converted to several GIS application formats through various frameworks. The data is stored in a relational database (PostgreSQL backend) which can be accessed, queried and edited by using a REpresentational State Transfer (REST) API, which uses HTTP GET, PUT and DELETE requests with XML payload. Data collection is acquired by GPS traces or by manual map modeling.

OSM uses a topological data structure with four data primitives: nodes (geographical positions stored as coordinates), ways (ordered lists of nodes which can represent a polygon), relations (multipurpose data structure that defines arbitrary relationships between 2 or more data elements which may be used to represent turn restrictions on roads), and tags (arbitrary metadata strings, which are mainly used to describe map objects). Each of these entities has a numeric identifier (called OSM ID) and a set of generic attributes defined by tags. For example, the natural tag describes geographical features which occur in Nature and has a wide set of values {bay, beach, … , wetland, wood} [16]. Further tags are used to specify time zones, currencies [17] and alike.

According to previous studies [18], the volunteered geographical information submitted to OSM is fairly accurate, with more than 80% of overlap between other specialized datasets, and often with more correct references in several countries around the world.

D. LinkedGeoData

LinkedGeoData (LGD) [2] is an effort to add a spatial dimension to the Semantic Web. It utilizes the information collected by the OpenSteetMap project and makes it available in Resource Description Frtamework (RDF) knowledge. The Semantic Web eases data and information integration by providing an infrastructure based on RDF ontologies, which are interactively transformed from OSM data. This procedure is believed to simplify real-life information integration that requires comprehensive background knowledge related to spatial features [17]. LinkedGeoData offers a flexible system for mapping data to RDF format, improved REST interface and direct interlinks to GeoNames and other geospatial data projects.

The data acquired from OSM is processed in different routes. The LGD Dump Module converts the OSM planet file into RDF and loads the data into a triple store. This data is then available via the static SPARQL endpoint. The LGD Live Sync Module monitors and loads change sets to the RDF level in order to update the triple store accordingly. The Osmosis is a community developed tool, which supports setting up such a database from a planet file and applying change sets to it. For data access, LGD offers downloads, a REST API interface, Linked Data and SPARQL endpoints [17]. The REST API provides limited query capabilities about all nodes of OSM.

LGD does not only contain linked data to OSM, but to other data services as well. It currently interlinks data with DBPedia and GeoNames [17], which is done on a per-class basis, where all instances of a set of classes of LGD are matched with all instances of a set of classes of other data sources using labels and spatial information. Since a vast amount of data is changed on OSM every few minutes, several filtering and synchronization procedures are used to ensure the data is kept relevant. Converting relational databases to RDF is also a significant area of research that LDG heavily depends on. These enhancements may further contribute to new semantic-spatial search engines to enable more linked data applications, such as geo-data syndication. However, certain limitations are still in the way of a more robust and scaling service, such as a lack of aggregating ontologies between OSM and the filtering system.

E. Discussion

Next, we discuss some aspects of online geodatabases that have been in the focus of the current research, namely, crowdsourcing, reliability and interoperability.

1) Crowdsourcing

A recent trend in neogeography has emerged by complementing information to geodatabases via several outside sources, namely social networks [19], shared media websites, and user's GPS contributions, known as Volunteered Geographical Information (VGI). The potential of crowdsourcing is that it can be a highly exploitable activity, in which participants are encouraged to contribute to an alleged greater good. For example, in Wikipedia, well over 99.8% of visitors to the site do not contribute anything [18]; yet, this does not deter the contributors from doing most of the work.

Since geospatial databases typically contain vast amounts of data, it is important to create mechanisms that can assist and verify the user-generated data submissions, which aim at improving internal data quality and resolving data conflicts between sources. These conflicts can be identified at the syntax (representation), structure (inconsistency), semantic (ambiguity) and cartographic (accuracy) levels [20].

Crowdsourcing in GIS has the following advantages:

- Make datasets accessible by non-proprietary languages and tools;
- Introduce formal semantics to make data machine learning possible, so the process can be automated between sources;
- Enrich existing data with other sources in a combined way, therefore increasing location precision;
- Allow cross-referencing and querying of multiple data sources simultaneously;

- Automatically detect on-going events and new places of interest from media and social networks.

However, there are several obstacles to this approach: data conflicts must be carefully managed, and proper conversion methods must be employed to ensure consistency (e.g., different names and tags for the same places in different sources), as well as other mechanisms to verify data integrity from user submissions (data validation). Outdated sources of information must also be filtered. A layer structure by Multi-Providers cRowd-Enhanced Geo linked Data (M-PREGeD) has been proposed [20] to improve data quality by involving the users in the selection process. This practice adopts the use of shared vocabularies and links to help increase semantic consistency and accessibility among sources of data.

In this case, results are stored in a local database and further enriched with matching Linked Geo Data nodes, which are verified and corrected by the Linked corrections layer, and only after that become available to the web applications layer. The user-generated matches are acquired by several applications, such as UrbanMatch [20], which is a game with a purpose for matching points of interest and photos, running on mobile geo-located services. The player is given a set of photos and is asked to match them with each location. The output is gathered from the players, verified and correlated and finally used to generate trusted links. A high number of similar answers are used to identify patterns of information and determine the accuracy of provided data.

This process can be defined in 3 layers: the Linked Geo Data layer which contains unprocessed data from several sources, the Linked Connections layer that is composed of all verification methods, such as the UrbanMatch game. Finally, this data is submitted to the web applications layer to the general audience.

Another process is done by analyzing metadata and tags from photos over different social media websites (such as Flickr and Twitter), by discovering associations between the media and the owner's geographical position, and is also possible to populate geo-sets with news data collected from other sources by employing extraction and retrieving systems [21].

These experiments not only help increasing spatial accuracy for existing places, but can also be used to discover new points of interest. Conflicts can be detected and resolved by automatic procedures, and expanding these techniques is an on-going process, especially when certain errors can occur [19] due to old or mislabeled data, incorrect semantic relationships, or simply by geographic evolution overtime. For example, there is a high correlation between places of interest like restaurants and a number of photos containing tags such as “food”, “dinner” and “eating” being uploaded on the same location. This quantitative evaluation consists of very specific statistical and data mining methods with several levels of precision. In recent studies [19], several new places in London were found with the help of these methods that were not present in GeoNames or Foursquare. Since different database providers are constructed in different ways, this procedure takes into account how different places are submitted and represented internally, and tries to aggregate similar results into a location vector, which creates a classifier that calculates the likelihood that a given location contains a place in particular. Closer examinations detected conflicts or mislabeled data such as misleading Tweets and incorrect or semantically ambiguous tags, and this allowed excluding potential outliers based on standard deviation in the dataset. Using this method, new places of worship, schools, shops, restaurants and graveyards were found that did not yet exist in LinkedGeoData and GeoNames. With finer iteration it becomes possible to detect new places of interest with higher precision, such as extending OpenStreetMap to indoor environments [22]. This approach aims at increasing the detail of geo data by adding new tags to inner structures, such as airports, museums or train stations by employing a 3D building ontology. Again, such data is likely to be supplied voluntarily by individuals at those places of interest. This kind of non-profit projects greatly enhances user participation.

2) Reliability

The proliferation of information sources as a result of networked computers has prompted significant changes in the amount, availability, and nature of geographic information. Among the more significant changes is the increasing amount of readily available Volunteered Geographic Information, which is produced by millions of users. However, many of these results are not provided by professionals, but by amateur users; therefore, they do not follow the common standards in terms of data collection, verification and use, this creates an issue which is very frequently discussed and debated in the context of crowdsourcing activities. Several studies [18] [23] [24] [25] have analyzed the systematic quality of VGI in great detail, and claim that for the most part results can be fairly accurate compared to other private or governmental entities, and represent the move from standardized data collection methods to data mining from available datasets.

Evaluating the quality of geographical information has received the attention of surveyors, cartographers and geographers many years ago, which have carefully deliberated a number of quality standards [18]:

- Lineage: the history of the dataset;
- Positional accuracy: how well the coordinate of an object is related to reality;
- Attribute accuracy: how well an object is represented with tag attributes;
- Logical consistency: the internal consistency of the dataset;
- Completeness: measuring the lack of data for a particular object;
- Semantic accuracy: making sure an object's representation is correctly interpreted;
- Temporal quality: the validity of changes in the database in relation to real-world changes and also the rate of updates.

Several methodologies were developed to quickly determine the data quality, from statistical comparison (using average standard deviation between different databases to calculate positional accuracy, object overlap percentages, tile boards, road length comparison among maps [25] and spatial

 ISBN: 978-1-61208-363-6

density among urban areas) to more empirical methods (visual density representations, road comparison). The results [25] state that there is an approximate overlap of 80% of motorway objects, roughly located within about 6m of distance recorded between OpenStreetMap and other proprietary map types in London, as well as less than 10% of total road length difference between OSM and TeleAtlas.

3) Interoperability

The growth of geospatial industry is stunted by the difficulty of reading and transforming suitable spatial data from different sources. Different databases have unalike conventions, classes, attributes and even entirely different semantic structures, which makes it difficult to interconnect and correlate data from different sources [26]. As billions of dollars are invested worldwide in the production of geospatial databases, it becomes imperative to find other alternatives for data translation and open, collaborative frameworks. As most GIS use their own proprietary format, translating this data into other formats can be a time consuming and inefficient process. Transformations between formats can also result in loss of information because different formats support different data types. Data translation also takes up a large amount of secondary storage and can be costly to develop. Other providers sometimes restrict access by repackaging products for a particular GIS.

There have been several solutions in the past to alleviate the issue, such as object oriented open frameworks and open architectures. Other approaches consist in interpreting other formats through the use of a universal language (such as, the Open Geospatial Datastore Interface (OGDI)). As the translation process is a long and difficult progress, OGDI is based on reading different geospatial data formats directly, without translation or conversion [10]. In essence, it works as a comprehension tool instead of translation. This allows for different formats to be queried from a uniform data structure, as well as transparent adjustment of coordinate systems and cartographic projections. Such data structures can be implemented via simple object oriented concepts, such as point features instantiating a coordinate, line features being composed of 2 or more coordinates in the same directional vector, and area features consisting of several line features forming a closed ring. This transient vector structure is the backbone of the retrieval functions, and allows for transparent adjustments and understanding of multiple coordinate systems. OGDI also uses an API that can validate parameters and sequences, transform coordinates and projections and provide an entry point to OGDI functions for each driver.

The multinational and multilevel initiatives, such as Open Geospatial Consortium (OGC) [27] and Defence Geospatial Information Working Group (DGIWG) [28], have been supporting and fostering the development of Digital Geospatial Information (DGI) standards that make spatial information more accessible to application developers.

III. SYMBOLIC CITY

In our previous work, we introduced our system architecture and the main structure of a symbolic world model. The world model we proposed is based on human mental models of space. It consists of objects, attributes, relations and relation attributes. The objects represent places and are further described by a set of attributes. Relations represent any kind of topological or other connection between places and also can be further described by a set of attributes. As such, our model can be visualized as a graph in which objects are nodes and relations are edges. A simple example is given in Figure 1.

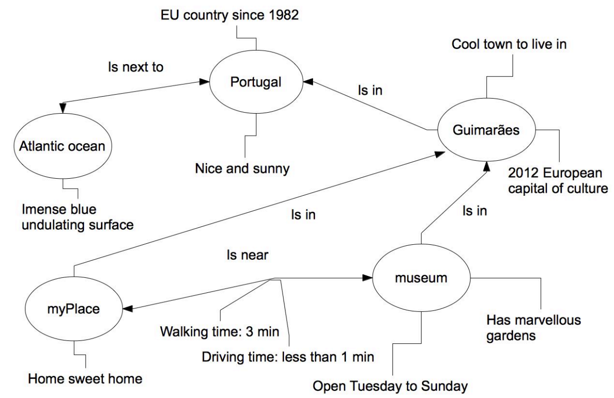

Figure 1. An exemple of a world model where places are nodes and relations are edges of a graph

Figure 2, in turn, shows how the symbolic model is integrated in the system architecture which supports different sources of data: manually introduced data, using a web application (SC Space Editor), automatically retrieved, by physical sensors and obtained on demand by software modules that communicate with online geodatabases and Place-Based Social Networks (PBSN).

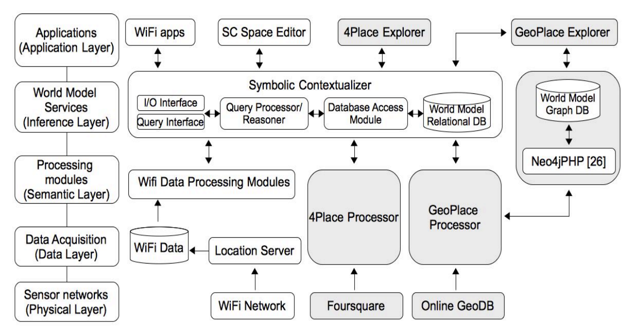

Figure 2. General system architecture

In the scope of the first type of interaction, the user can insert, modify, remove and query all the objects represented in the world model. In the second case, a set of processing modules was developed to extract data from WiFi networks, process them and insert them automatically in the model. Finally, the third type of interaction with the symbolic model is based on data from online geodatabases and PBSN, namely Foursquare. This type of data feed can be considered semi-automatic, as it requires some user intervention, as we will further explain.

 ISBN: 978-1-61208-363-6

We envision an application called Symbolic City that is based upon our world model and provides a new way of looking at places of interest in a city, as shown in Figure 3.

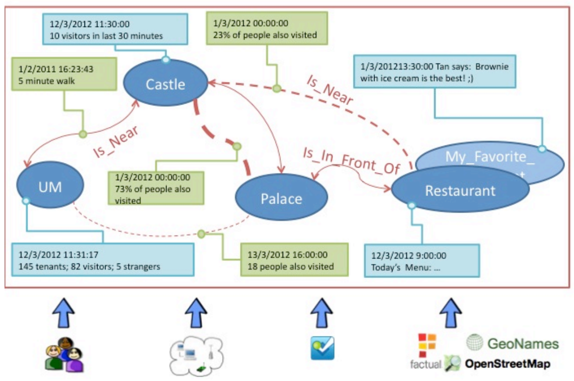

Figure 3. Symbolic City service concept

In order to achieve this kind of data crossing, we propose several sources of data and a set of processing modules that feed the model and allow for inference of place attributes that change with time, the creation of relations between places and user world model customization. The editing application was object of our previous work as well as the modules corresponding to WiFi data analysis. In the following subsections we describe the two remaining modules.

A. 4Place Explorer

Location integration in social networking applications provides means for people to share in real time places they or their friends and acquaintances visit as well as look for tips, comments and ranks of Points Of Interest (POIs) or places they intend to visit. Foursquare, Google+ and Facebook users are able to share their present location through the check-in mechanisms and comments about these places. Currently, these three applications are most commonly used in the scope of location sharing.

4Place Explorer is a PHP module we implemented to allow for the integration of the available data from Foursquare and Google Maps (for the purpose of route finding between two places) with the world model data. It uses Foursquare API. OAuth standard is used as it provides the access to most of the API functions with a user account.

In the current version of this module, the user searches for a place by its name. Data about that place are fetched from Foursquare in JSON format. The user chooses the matching place and a query in the symbolic model is made to check if that place already exists. If it exists, only the new attributes and new relations are inserted in the model, if there is any. If it does not exist, a new object is created with a set of attributes and a set of relations. The attributes that are extracted are the following: name, category, address, coordinates, city and country. The current number of *check-ins* may act as an indicator of how many people are visiting a place and it also may be added to the model. Currently, three types of relations are created automatically: Is_In, between an object and a city or a country; Is_Near, based on the distance between two places, calculated in a given range from the found object; and Is_Accessible_From, if a route can be obtained from Google Maps.

The purpose of this module is to provide a new perspective over the places that are represented in the two sources of geographical data, Foursquare and symbolic world model. The implemented mechanisms allow for automatic creation of new objects in the model as well as connecting objects with relevant topological relations.

The existence of relations transforms a set of dispersed places to a graph of interrelated places, providing not only proximity of points of interest, but also inviting the visitor to nearby locations and contributing to a better publicity, be it for tourism or for commercial purposes.

B. GeoPlace Explorer

Current version of the GeoPlace Explorer application allows for creating new places manually and searching for a place by name in Factual, GeoNames and OpenStreetMap geodatabases. After the search results are shown, the user choses the object to be created in the database. Objects have default attributes such as name, category, latitude and longitude. The user that creates an object becomes its author. The online geodatabase source from which the object was retrieved is also registered. An object may have other free attributes specified during its creation. Objects belong to one of six basic categories (city, island, building, country, road and ocean) and can have relationships (with corresponding attributes) to one or more objects. For example, the object of category "city", named Funchal can be related (relation "Is_In") to the object called Madeira of category "island", which, in turn, is related to the object Portugal, of category "country".

Our database is modeled in Neo4j [29], one of the most popular graph database models. The object creation and manipulation is done by a series of PHP files that interact with the database by the means of the Neo4j client library called Neo4jPHP [30] using Cypher queries.

The purpose of this module is twofold: to learn and test a graph database as a possible solution for our world model repository; and to analyze APIs and data about POIs from three different online geodatabases. So far, the main findings were that graph databases are farther more flexible and scalable than the relational databases as they do not require a rigid predefined structure and that using the three studied online geodatabases is advantageous as the existing data may be complementary.

IV. CONCLUSION AND FUTURE WORK

In this paper, we presented an overview of the most popular online geodatabases and discussed their main characteristics and strengths. Crowdsourcing has shown to be the most significant way of maintenance of large spatial data repositories, be it directly, as in OSM, be it through social networking, as in Foursquare. We introduced a symbolic world model as a repository of spatial data created manually by the user through a web application as an image of their mental model of space, and automatically by a set of processing modules that interact with sensors, place-based social networking applications and online geodatabases.

The envisioned Symbolic City application is related with spotting not only the POIs in a city, but also the relations that exist between POIs and that are inferred automatically by the developed modules. We believe that the users may get more details about POIs of their choice by adding them to their own personal models by combining several different sources of data. We expect to achieve this as the integration of the SC Space Editor, the 4Place Explorer and the Geo Explorer may follow in the near future.

However, there are still several challenges that we face in our present stage of research. They are related to the social and collaborative aspect of our application, with the choice of a database implementation (transition from a relational to a graph database or a hybrid solution), system distribution over several servers, and so on. There are already some promising solutions for each of these challenges, so in the near future we expect to obtain some new and exciting results.

ACKOWLEDGMENT

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: PEst-OE/EEI/UI0319/2014.

REFERENCES

[1] "GeoNames." [Online]. Available: http://www.geonames.org/. [Accessed: 18-Jun-2013].

[2] "linkedgeodata.org : About." [Online]. Available: http://linkedgeodata.org/About. [Accessed: 15-May-2014].

[3] "Factual | Home - The Location Platform Enabling Mobile Personalization." [Online]. Available: http://www.factual.com/. [Accessed: 15-May-2014].

[4] "Facebook Asks Users to Clean Up Its Location Database with Places Editor and Favorite Places," *Inside Facebook*. [Online]. Available: http://www.insidefacebook.com/2011/07/01/favorite-places-editor-location-database/. [Accessed: 08-Jul-2013].

[5] K. Baras and A. Moreira, "Groups and Frequent Visitors Shaping the Space Dynamics," in *Smart Spaces and Next Generation Wired/Wireless Networking*, vol. 6869, S. Balandin, Y. Koucheryavy, and H. Hu, Eds. Springer Berlin / Heidelberg, 2011, pp. 111–122.

[6] B. Tversky, "Remembering Spaces," in *The Oxford handbook of memory*, E. Tulving and F. I. M. Craik, Eds. New York, NY, US: Oxford University Press, 2000, pp. 363–378.

[7] K. Baras, "Dynamic world model for context-aware environments," Ph.D. Dissertation, DSI, Universidade do Minho, Guimarães, Portugal, 2012.

[8] A. Markowetz, T. Brinkhoff, and B. Seeger, "Exploiting the internet as a geospatial database," presented at the Workshop on Next Generation Geospatial Information, 2003.

[9] "OpenStreetMap." [Online]. Available: http://www.openstreetmap.org/about. [Accessed: 15-May-2014].

[10] M. Schmidt and P. Weiser, "Web Mapping Services: Development and Trends," in *Online Maps with APIs and WebServices*, M. P. Peterson, Ed. Springer Berlin Heidelberg, 2012, pp. 13–21.

[11] "How Google Builds Its Maps—and What It Means for the Future of Everything," *The Atlantic*, 06-Sep-2012. [Online]. Available: http://www.theatlantic.com/technology/archive/2012/09/how-google-builds-its-maps-and-what-it-means-for-the-future-of-everything/261913/. [Accessed: 25-Jul-2013].

[12] F. Derby, "Coordinates : A resource on positioning, navigation and beyond » Blog Archive » Geospatial databases and the evolving role of the surveyor," 21-Mar-2014. [Online]. Available: http://mycoordinates.org/geospatial-databases-and-the-evolving-role-of-the-surveyor/. [Accessed: 21-Mar-2014].

[13] National Research Council, *IT Roadmap to a Geospatial Future*. Washington DC: National Academies Press, 2003.

[14] V. Maltese and F. Farazi, "A semantic schema for GeoNames," Information Engineering and Computer Science, University of Trento, Departmental Technical Report DISI-13-004, 2013.

[15] "2013 OpenStreetMap Data Report," 21-Mar-2014. [Online]. Available: https://www.mapbox.com/osm-data-report/. [Accessed: 21-Mar-2014].

[16] P. Mooney, P. Corcoran, and A. C. Winstanley, "A study of data representation of natural features in openstreetmap," presented at the 6th International GIScience Conference, 2010, vol. 150.

[17] C. Stadler, J. Lehmann, K. Höffner, and S. Auer, "Linkedgeodata: A core for a web of spatial open data," *Semantic Web*, vol. 3, no. 4, pp. 333–354, 2012.

[18] M. Haklay, "How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets," *Environ. Plan. B Plan. Des.*, vol. 37, no. 4, pp. 682–703, 2010.

[19] S. Van Canneyt, O. Van Laere, S. Schockaert, and B. Dhoedt, "Using Social Media to Find Places of Interest: A Case Study," in *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information*, New York, NY, USA, 2012, pp. 2–8.

[20] R. Karam and M. Melchiori, "Improving geo-spatial linked data with the wisdom of the crowds," in *EDBT '13 Proceedings of the Joint EDBT/ICDT 2013 Workshops*, 2013, pp. 68–74.

[21] P. López and J. César, "Geospatial database generation from digital newspapers: use case for risk and disaster domains.," MSc, ISEGI, Universidade Nova de Lisboa, Lisbon, 2010.

[22] M. Goetz and A. Zipf, "Extending openStreetMap to indoor environments: bringing volunteered geographic information to the next level," *Proc. Urban Reg. Data Manag. Udms Annu. 2011*, pp. 47–58, 2011.

[23] A. J. Flanagin and M. J. Metzger, "The credibility of volunteered geographic information," *GeoJournal*, vol. 72, no. 3–4, pp. 137–148, Aug. 2008.

[24] "The Citizen Science Landscape: From Volunteers to Citizen Sensors and Beyond," 21-Mar-2014. [Online]. Available: http://www.hindawi.com/journals/ijz/2012/349630/. [Accessed: 21-Mar-2014].

[25] D. Zielstra and A. Zipf, "A comparative study of proprietary geodata and volunteered geographic information for Germany," presented at the 3th AGILE international conference on geographic information science, Guimarães, Portugal, 2010, vol. 2010.

[26] S. Hahmann and D. Burghardt, "Connecting linkedgeodata and geonames in the spatial semantic web," presented at the 6th International GIScience Conference, 2010.

[27] Open Geospatial Consortium, "Open Geospatial Consortium | OGC(R)," *Open Geospatial Consortium | OGC(R)*, 2014. [Online]. Available: http://www.opengeospatial.org/. [Accessed: 16-Jun-2014].

[28] DGIWG, "DGIWG Home Page," *Defence Geospatial Information Working Group*, n.d. [Online]. Available: https://www.dgiwg.org/dgiwg/index.htm. [Accessed: 16-Jun-2014].

[29] Neo Technology, Inc., "Neo4j - The World's Leading Graph Database," *Neo4j - The World's Leading Graph Database*, 2014. [Online]. Available: http://www.neo4j.org/. [Accessed: 16-Jun-2014].

[30] J. Adell, "Everyman Software: Development Setup for Neo4j and PHP: Part 1," *Everyman Software Musings on software development, open source, PHP and current projects.*, 05-Nov-2011.

 ISBN: 978-1-61208-363-6