

Designing SDI4Apps POI Base

Otakar Čerba¹, Tomáš Mildorf¹, Raitis Berzins²

¹ University of West Bohemia, Univerzitní 8, 306 14 Plzeň, Czech Republic
{cerba, mildorf}@kma.zcu.cz

² Baltic Open Solutions Center, Krišjāņa Barona iela 32-7,
Rīga, LV-1011, Latvija
raitisbe@gmail.com

Abstract. The SDI4Apps project has collected a large number of points of interest (POIs). This data set represents a seamless and open resource of POIs in Europe. Its principal target is to provide information for cycling as Linked data together with other data set containing road network. The POIs, which will be available for other users for download, search and reuse, will be helpful for other applications in tourism as well. The article presents the data model for POIs and harmonization of external data sources into this data model. The current version of the SDI4Apps POI data set includes a harmonized combination of selected OpenStreetMap data, experimental ontologies and local data. A short comparison of the SDI4Apps POIs with the OpenPOIs data set is presented.

Keywords: Point of interest, Linked Data, data model, SDI4Apps, data set, spatial data modeling.1 Introduction

1 Introduction

SDI4Apps¹ is an EU-funded project (European Union's ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme) coordinated by the University of West Bohemia² in Plzen, Czech Republic. SDI4Apps seeks to build a cloud-based framework with open APIs (application programming interfaces) for data integration focusing on the development of six pilot applications. The project draws along the lines of INSPIRE (INfrastructure for SPatial InfoRmation in Europe), Copernicus and GEOSS (Global Earth Observation System of Systems). The SDI4Apps development process started with data integration and harmonization, including semantic annotation and Linked Data interconnection. Data are collected

¹ <http://sdi4apps.eu/>

² <http://www.zcu.cz/>

based on requirements of the six pilot activities, which represent the first users, testers and feedback providers of the whole SDI4Apps solution.

This article describes the SDI4Apps Point of Interest (SPOI) dataset as a specific set of POIs which are useful for potential customers of applications developed in the SDI4Apps project, above all in the Open Smart Tourist Data pilot. “POI provide an essential data source for a wide range of location-based applications. Having emerged from in-car navigation systems the classic POI are often linked to an address and relate to businesses such as petrol stations, garages, shopping centers, or common sense information, such as the church, police station or hospital in a city.” (Andrae et al., 2011). Various POI datasets are implemented into popular applications for traveling and tourism such as Trip Advisor or navigation tools such as Waze. There are also two US Patents mentioning the role of POIs in on-line advertisement (Jakobson & Rueben, 2013a, Jakobson & Rueben, 2013b).

There are a lot of resources providing POI data. First of all, it is necessary to mention OpenPOIs by Open Geospatial Consortium. This database contains more than 9 million POIs, which are available through an API. Other POI data are offered by various web pages such as POIplaza³, POI download⁴, GPS Data Team⁵, Pocket GPS World⁶ or the POI service provided by Flemish government⁷. Data are also provided by several producers of navigation tools. These resources contain various types of POIs and enable to download data in different formats (usually in a format that can be processed by navigation tools).

The SPOI data set is created as a combination of global data (selected points from OpenStreetMap) and local data provided by the SDI4Apps partners or data available on the web. The final version will represent an open and seamless solution which will be able to be “a data fuel” for location-based and navigation services and applications. The added value of the SDI4Apps approach consists in implementation of linked data. The current version contains several links to external resources (see the part “SPOI as Linked Data”). Interconnections to other data will be added, including transformation of all used code list into RDF (Resource Description Framework) vocabularies.

There are several disadvantages of contemporary POIs datasets (such as those mentioned above), which prevent their integration and further re-use. These disadvantages include:

- using proprietary formats or specific formats for geographic information systems,
- download based on topics or geographical regions (for example countries),

³ <http://poi plaza.com/>

⁴ <http://www.downloadpoi.com/>

⁵ <https://www.gps-data-team.com/>

⁶ <http://www.pocketgpsworld.com/>

⁷ <http://poi.api.geopunt.be/>

- common absence of standardized services or querying,
- charge for data.

The main goal of this article is to introduce the SPOI base, above all the development process and data model. The SPOI data are compared with the similar solution OpenPOIs in order to check the compatibility between both POI datasets for further reuse in various applications.

The article is structured as follows. Section 2 describes the used methodology including fundamental pillars of the design and development of the SPOI base. This includes theoretical backgrounds as well as inspiring data sets and models (as the overview of state-of-the-art). The methodology also contains a short description of the comparison of the SPOI base and OpenPOIs. Section 3 presents the SPOI data set, its data model and relation to the Linked Data approach. Section 4 includes the comparison of the SPOI and OpenPOIs datasets. This section shows the potential of combining both data resources. The last section (except Conclusions) includes a discussion on further steps of the SPOI development.

2 Methodology

The reason for the development of another dataset of POIs emerged from the needs of users (tourists, tourist service providers as well as developers of applications focused on tourism). There was a lack a complex set of POIs which would not be territory specific (for example limited to particular countries, regions or national parks), would be open and not limited to one data resource (usually OpenStreetMap). The SDI4Apps team composed above all of experts from the Czech Republic and Latvia developed a seamless open database of POIs which will be distributed as 5-star Linked Open Data (Berners-Lee, 2009) to be accessible for all users.

Even though the data modeling of POIs (as features with simple point geometry, identifier and several descriptive attributes) seems to be very trivial, authors did an extensive research of existing data models and literature. The development of the SPOI data model was based on seven fundamental pillars:

1. Classical studies and books focused on spatial (or geographical) data modeling such as Goodchild (1992), Shekhar et al. (1997), Longley et al. (2001) or Tomlinson (2007). These resources gave a basic framework of SPOI data model.
2. Because RDF has to be the principal format to store data, also several researches dealing with publication and modeling of spatial data as RDF triples (for example Auer et al., 2009, Janowicz et al., 2012 or Kritikos et al., 2013) were taken into consideration and implemented.

3. General principles of development Linked Data as they are published in Bizer et al. (2008), Heath et al. (2008), Bizer et al. (2009) or Hausenblas (2009). There is also a lot of publications focused on Linked Data on the geographical domain such as Auer & Lehmann (2009), Atemezing & Troncy (2012) or Kuhn et al. (2014).
4. The data model of POI as it is published in the W3C Editor's Draft Points of Interest Core (Hill & Womer, 2012; this document was originally created W3C Points of Interest Working Group that was transformed into OGC as Points of Interest Standards Working Group) as well as in the presentation Framing a Geo Strategy for the Web with Points-Of-Interest Data by R. Singh (2012).
5. Data models of existing POI datasets (for example POIplaza or POI download).
6. Existing standards, formats and vocabularies such as RDF, RDFS, SKOS, OWL, FOAF, GeoSPARQL or WGS84 Geo Positioning.
7. Experiences of data modeling from existing solutions and projects such LinkedGeoData, DBpedia, GeoNames.org or SmartOpenData.

The reasons for selection of particular classification systems, coordinate systems and other parts of the data model are explained in the next section.

The POI data model is open and flexible. The essential core of the model (Id, coordinates, label and categorization) was extended by several attributes which are integral components of some original data and could be helpful for tourist purposes (for example contact information, opening hours or accessibility for handicapped visitors).

The contemporary version of the SPOI base is populated by XSLT templates (several examples of using XSLT transformations in spatial data domain are published in Čerba, 2010 or Čerba & Čepický, 2012). The XSLT template contains procedure of harmonization, data models' mapping, including transformation rules of classification systems.

In order to prove the concept of the SPOI base, a comparison with the OpenPOIs data set as the main sample of global POI database was realized. The first part of the comparison contains basic properties of both data sets such as coverage, number of POIs or output formats. The second part tests five small areas in Europe and its coverage by POIs in both databases. The size of the each area is $0,02^\circ \times 0,02^\circ$ to satisfy the limitation of OpenPOIs. The OpenPOIs API is able to provide in maximum 100 POIs. Ten various types of landscape were chosen as areas of interest (for example city important for tourism, mountains, coast, industrial area or countryside). Also various European countries (for example Czech Republic, France, Italy, Latvia,

Poland) were selected to limit local differences. Numbers of POIs were gained by XSLT template for filtering data (SPOI) and custom API (OpenPOIs).

3 SDI4apps POI Base

3.1 Basic description

Table 1. SPOI – basic information.

Property	Description
Data amount	3 292 230 POIs
File size	3,1 GB
Coverage	Europe (45 countries)
Data sources	OpenStreetMap Local data from the Posumavi region (Czech Republic) Experimental ontologies developed at the University of West Bohemia (Czech Republic) – European ski resorts and religious monuments in Rome
POI classification	SPOI contains nine fundamental classes adopted from the data model used for data of the Waze navigation tool.
Storage	XML file Virtuoso
Publication	Virtuoso SPARQL endpoint ⁸ Map application Smart Tourist Data (Geoportal SDI4apps)
Links	Several POIs are linked to DBpedia and GeoNames.org. There are also DBpedia and GeoNames.org links to particular countries containing POIs. The main classification of POIs is accessible through URI.

Table 1 presents basic information about the set of the POIs developed in the Open Smart Tourist Data pilot application of the SDI4Apps project.

⁸ <http://ha.isaf2014.info:8890/sparql>

3.2 SPOI as Linked Data

SPOI corresponds with 5-star rating system of Linked Open Data published by Berners-Lee (2009) and described in Janowicz et al. (2014). Table 2 shows how particular criteria are satisfied by the SPOI data.

Table 2. SPOI & 5-star rating system of Linked Open Data.

Stars	Description	SPOI
1	Data is available on the Web under an open license.	Data are provided to download on the Web through the SPARQL endpoint. The data will be provided under the Open Database License (ODbL).
2	Data are available as a structured data.	SPARQL endpoint is able to provide data in many structured formats, including JSON, XML, CSV or various serialization of RDF.
3	Data uses a non-proprietary format.	Majority of output format offering in Virtuoso SPARQL endpoint are classed as non-proprietary formats.
4	Particular objects have URI as identifier.	Data uses unique identifier based on URI based on http://www.sdi4apps.eu/poi .
5	Data is linked to another data.	Several object are linked by properties <code>skos:exactMatch</code> and <code>owl:sameAs</code> to equivalent elements in DBpedia and GeoNames.org. All objects are interconnected via topological property <code>sfWithin</code> to relevant countries as they are expressed in DBpedia and GeoNames.org. The main classification of POIs is accessible through URI.

3.3 Data model

The contemporary version of the SPOI data model (June 2015, Fig. 1) has seven basic components:

1. Identification – each POI is identified by unique ID expressed as URI. Original ID (URI of the product and a unique code generated by XSLT script) was replaced by more readable form providing some information such as country and category. The new identifier is composed of URI (<http://www.sdi4apps.eu/poi>), ISO 3166-1 alpha-2 country code, category of

POI according Waze navigation data and code (generated randomly by the XSLT script).

2. Description – each POI is described by a label (name). In several cases, there are more labels differentiated by the xml:lang attribute. POIs can contain a text description if it is available.
3. Geometry / Localization – each POI is localized by two coordinates (latitude and longitude) of World Geodetic System (WGS) 84. WGS84 represents the most used, respected and universal system, which is usually transformable to local systems and cartographic projections. Coordinates are published according to Basic Geo (WGS84 lat/long) Vocabulary (Brickley, 2006).
4. Classification – categorization is realized through three various parameters – classification based on GPS-based geographical navigation Waze, which is primary, mandatory and used for visualization on the SDI4apps geoportal. The classification system used in Waze is quite short, clear and simple to visualize as well as differentiate, because it contain 10 well-defined categories. Since majority of data originate in OpenStreetMap, two types of classification from Open StreetMap are used. The authors tested other nomenclatures used in various products (data, services, applications) such as Trip Advisor, Yelp!, USGS Geographic Names Information System or Ordnance Survey POI classification scheme, but the Waze scheme is the most appropriate to purposes of POI database developed in the SDI4Apps project. Mapping rules between the Waze nomenclature, the OpenStreetMap classification and categories used in other source data are kept in the transformation XSLT file.
5. Contact information – several POIs contain contact information such as address, e-mail, homepage, fax or phone number.
6. Common information – currently there are only two types of this type of information (opening hours and access). This information is available only for data from the Posumavi region.
7. Links – all POIs include one or more of three types of links to external data – links to external non-linked data resources such as Wikipedia, Wolfram|Alpha or raster maps, links to an equivalent object in DBpedia or GeoNames.org, links to countries (in DBpedia and GeoNames.org) containing the POI. The last type of links is mandatory for each object.

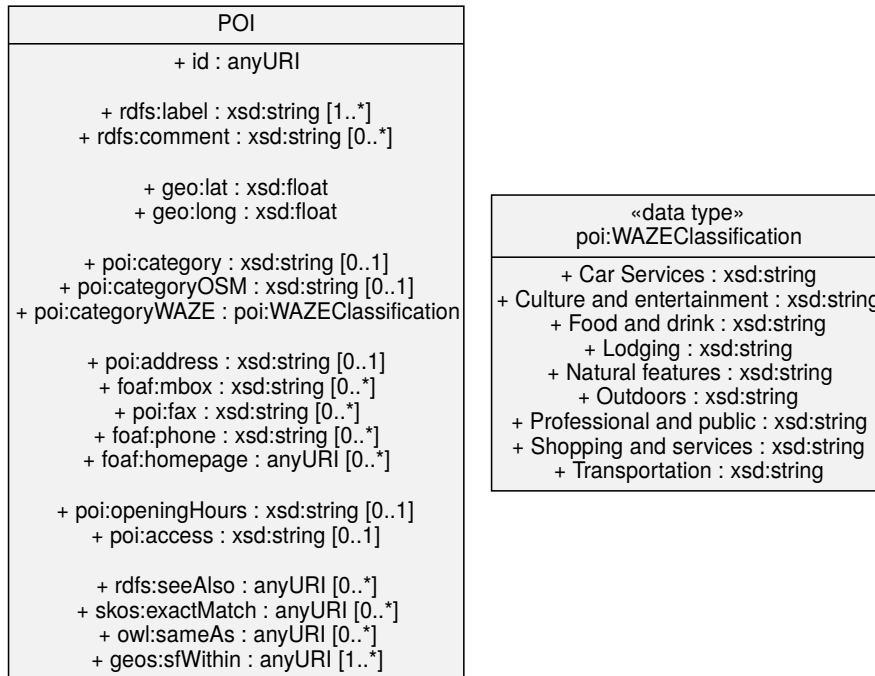


Fig. 1. SPOI – data model.

4 Comparison of SPOI and OpenPOIs

To evaluate the concept of the SPOI base, a short comparison with a similar data set (OpenPOIs) was performed. This part is divided into two parts – comparison of common characteristics and monitoring of POIs in selected areas.

4.1 Common characteristics

Table 3 shows basic properties of both data sets. Information on OpenPOIs was extracted from the OpenPOIs' homepage⁹ and presentation Framing a Geo Strategy for the Web with Points-Of-Interest Data (Singh, 2012).

⁹ <http://openpois.net/>

Table 3. SPOI & OpenPOIs – basic characteristics.

Property	SPOI	OpenPOIs
Number of POIs	> 3.2 millions	> 9.5 millions
Coverage	Europe	World
Main sources of data	OpenStreetMap	GeoNames, DBpedia (these resources are mentioned in Singh, 2012, a short survey of data demonstrated that many objects originated from OpenStreetMap)
Ways of data providing	SPARQL endpoint	Custom API, WFS
Output data formats	Formats provided by Virtuoso tool (RDF, JSON, CSV, Javascript...)	XML, JSON, microdata, RDF

Table 3 shows two main relations between both POI data sets. Both are based on similar original data (OpenStreetMap). This fact will be evident from Table 4 comparing numbers of POIs in selected areas. The second similarity is connected with standardized solutions to provide data. OpenPOIs prefers standards of OGC, because this organization maintains OpenPOIs as well as Web Feature Service. SPOI uses SPARQL endpoint to be compliant with Linked Data and RDF solutions. Output formats of both sets are similar.

Also data models are comparable. Both data models contains basic components such as identifiers, location (points; but both products deal with addresses and OpenPOIs also with relationships to other POIs), labels, description, categorization and links. OpenPOIs offers metadata items. SPOI contains more contact information and data important for issues of tourism.

The most noticeable difference is evident from data. OpenPOIs data are just copied from original resources, while SPOI data are harmonized to the uniform data model. Therefore all SPOI features use the same classification in comparison with SPOI. This fact can complicate potential combination of both data sets, but it can be treated with similar transformation rules as they are applied to import external data. Moreover, mapping between the OpenStreetMap classification of POI and the Waze nomenclature, which is used as primary classifier in SPOI, is defined in the contemporary version of XSLT styles.

4.2 POIs in selected areas

The comparison of quantity of POIs in both datasets (Table 4) was realized in ten European localities. These areas (0,02° x 0,02°) were selected to cover various types of landscape (for example rural area, industrial area, large city or mountains) as well as different countries evenly distributed over the whole continent.

Table 4. Number of POIs in selected areas.

Area	SPOI	OpenPOIs
Seaside resort (Croatia)	7	4
Submontane area (Czech republic)	1	0
Mountains (France)	1	1
Rural area (Germany)	28	28
Historical site (Greece)	9	10
Large city (Italy)	57	60
Coast (Latvia)	0	0
Small towns and villages (Netherlands)	6	8
Sport center (Norway)	46	41
Industrial area (Poland)	54	57

Even though the total number of POIs in selected sample areas (Table 4) is equal (209), there are evident several interesting knowledge (which can utilized to improve and extend SPOI data, because complete integration of OpenStreetMap data is not finished yet):

1. OpenPOIs dataset contains more POIs in urbanized regions.
2. Results in sparsely populated area is very similar.
3. In localities important for tourism SPOI data evinces better results.
4. SPOI and OpenPOIs show similar results in post-communist countries.
5. There are minor differences between North and South Europe.

The Table 1 shows that both POI data sets are quite similar. It is evident not only from the number of POI, but also from the similar content. For example the sample from France containing just one same POI in both resources. This fact just support a challenge of joining of SPOI and OpenPOIs to get large POI database. In this case it is necessary to solve redundant feature in both datasets.

This test is just initial. It will be repeated in other areas to find out potential random errors and prove hypothesis mentioned in previous list. Also a graphical visualization

(a heat map) and comparison of content (not only number of POIs) will be realized to compare both data sets.

5 Discussion & future steps

The contemporary version of the SPOI base is useful, but the base as well as its model haven't been completed yet. The developers together with the SDI4Apps project and other users are discussing many proposed changes and improvements. They could be divided into two groups – (1) modifications of the data model and (2) other further steps related to populating, visualization or maintenance.

The possible changes of the SPOI data model include implementations of:

- a secondary identifier based on name(s) of features to make URIs more readable.
- persistent identifiers being stable during data updating.
- the form of coding coordinates as it is defined in the GeoSPARQL standard (Perry & Herring, 2011) to support exploitation of the GeoSPARQL querying.
- a property for preferred label (for example skos:prefLabel), because the SPOI base contains more than one labels (in one language) for several features.
- transformations of classifications to RDF structure to be re-usable in other data and applications.
- changes of string values (for example addresses or opening hours) to several semantically rich values, for example based on INSPIRE specifications (for example INSPIRE Data Specification on Addresses) or ISA Core Location Vocabulary for addresses.
- new attributes important for tourism.

Other further steps are related to population of the SPOI base (searching of new data resources and its processing, removing errors and shortcomings in data, massive and automated adding links to other resources), refining data (eliminations of duplicities), providing data (export to other formats that are supporting by navigations tools, improvements of map portal, generalization), updating (questions of persistent URIs or processing of changes in source data) and improvements of a presentation of the product (social media). A detail description of these steps is not the subject of this article and the final list will change according to user requirements.

6 Conclusions

There are many ways how to describe the SDI4Apps POI data set (for example quality of data or maintenance and updating). With respect to the limits of the conference proceedings it is not possible to mention all question or problems connected to SPOI.

This paper introduces the data set of Points of interest developed in the SDI4Apps project. This data set is the seamless and open resource of POIs that will be available for other users to download, search or use in applications and services. The data model of SPOI comes from review of literature, existing data (for example OpenPOIs), recommendations of W3C and OGC and user requirements. The current version of the data set has been created as a harmonized combination of selected OpenStreetMap data, experimental ontologies developed in the Section of Geomatics of the University of West Bohemia and local data provided by the Uhlava region (Czech Republic). The transformation was realized by XSLT templates. Data are stored in the Virtuoso tool as RDF triples. SPOI is published via SPARQL endpoint which enables comfortable, efficient and standardized querying of data.

The document also contains the short comparison of SPOI with other respected data set of POIs (OpenPOIs). The results of the comparison show common data resources (above all OpenStreetMap), similar approach to data modeling and contrarily various approaches to data harmonization, storing and provision. The acquired information could be useful in case of development of mix of both data set or in case of mutual exchange of data.

The added value of the SDI4Apps approach in comparison to other similar solutions consists in implementation of linked data, using of standardized and respected datatype properties and development of the completely harmonized data set with uniform data model and common classification (not only a copy of original resources).

The authors believe that the selected approach to develop an open data base of POIs is promising, because

- implementation of many external data resource can provide a multi-level view on POIs, including corrections of shortcoming and gaps,
- Linked data enables more efficient way how to combine and re-use data,
- open data can generate an interesting business effect such local advertising or development applications.

The authors welcome other remarks and comments how to improve the SPOI data set, its model, content as well as interconnection to other data.

References

1. Andrae, S., Erlacher, C., Paulus, G., Gruber, G., Gschliesser, H., Moser, P., Sabitzer, K. & Kiechle, G. (2011). OpenPOI—Developing a web-based portal with high school students to collaboratively collect and share points-of-interest data. *Learning with geoinformation*, 66-69.
2. Atemezing, G. A., & Troncy, R. (2012). Comparing vocabularies for representing geographical features and their geometry. In *Terra Cognita 2012 Workshop* (p. 3).
3. Auer, S., & Lehmann, J. (2009). LinkedGeoData—Collaboratively Created Geo-Information for the Semantic Web. *Semantic Web Challenge, ISWC*.
4. Auer, S., Lehmann, J., & Hellmann, S. (2009). Linkedgeodata: Adding a spatial dimension to the web of data (pp. 731-746). Springer Berlin Heidelberg.
5. Berners-Lee, T. (2009). *Linked Data. Design Issues*. W3C.
6. Bizer, C., Heath, T., & Berners-Lee, T. (2008). Linked data: Principles and state of the art. In *World Wide Web Conference*.
7. Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. *Semantic Services, Interoperability and Web Applications: Emerging Concepts*, 205-227.
8. Brickley, D. (2006). Basic Geo (WGS84 lat/long) Vocabulary, World Wide Web Consortium.
9. Čerba, O. (2010). XSLT Templates for Thematic Maps (pp. 181-192). Springer Berlin Heidelberg.
10. Čerba, O., & Čepický, J. (2012). Web Services for Thematic Maps. In *Online Maps with APIs and WebServices* (pp. 141-155). Springer Berlin Heidelberg.
11. Goodchild, M. F. (1992). Geographical data modeling. *Computers & Geosciences*, 18(4), 401-408.
12. Hausenblas, M. (2009). Exploiting linked data to build web applications. *IEEE Internet Computing*, (4), 68-73.
13. Heath, T., Hausenblas, M., Bizer, C., Cyganiak, R., & Hartig, O. (2008). How to publish linked data on the web. In *Tutorial in the 7th International Semantic Web Conference, Karlsruhe, Germany*.
14. Hill, A. & Womer, M. (2012). Point of Interest Core. W3C Editor's Draft.
15. Jakobson, G., & Rueben, S. (2013a). U.S. Patent Application 13/986,744.
16. Jakobson, G., & Rueben, S. (2013b). U.S. Patent Application 13/987,075.
17. Janowicz, K., Hitzler, P., Adams, B., Kolas, D., & Vardeman, C. (2014). Five stars of Linked Data vocabulary use. *Semantic Web*, 5(3), 173-176.
18. Janowicz, K., Scheider, S., Pehle, T., & Hart, G. (2012). Geospatial semantics and linked spatiotemporal data-Past, present, and future. *Semantic Web*, 3(4), 321-332.
19. Kritikos, K., Rousakis, Y., & Kotzinos, D. (2013). Linked open GeoData management in the cloud. In *Proceedings of the 2nd International Workshop on Open Data* (p. 3). ACM.
20. Kuhn, W., Kauppinen, T., & Janowicz, K. (2014). Linked data-A paradigm shift for geographic information science. In *Geographic Information Science* (pp. 173-186). Springer International Publishing.

21. Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2001). *Geographic information system and Science*. England: John Wiley & Sons, Ltd.
22. Perry, M., & Herring, J. (2011). OGC GeoSPARQL-A geographic query language for RDF data. OGC Implementation Standard, ref: OGC.
23. Singh, R. (2012). Framing a Geo Strategy for the Web with Points-Of-Interest Data. Invited presentation in the 5th International Terra Cognita Workshop 2012 In Conjunction with the 11th International Semantic Web Conference, Boston, USA.
24. Shekhar, S., Coyle, M., Goyal, B., Liu, D. R., & Sarkar, S. (1997). Data models in geographic information systems. *Communications of the ACM*, 40(4), 103-111.
25. Tomlinson, R. F. (2007). *Thinking about GIS: geographic information system planning for managers*. ESRI, Inc..