Towards Spatial Knowledge Infrastructure (SKI): Technological Understanding

Morteza Omidipoor
Tehran University
Department of Remote Sensing and GIS
Tehran, Iran
omidipoor@ut.ac.ir

Ara Toomanian
Tehran University Department of Remote Sensing and GIS
Tehran, Iran
a.toomanian@ut.ac.ir

Najmeh Neisany Samani
Tehran University Department of Remote Sensing and GIS
Tehran, Iran
nneysani@ut.ac.ir

Abstract

Nowadays, a growing attention has been paid to extracting knowledge from voluminous and distributed spatial data. In this regard, a number of spatial data mining methods and algorithms have been developed, but there are a number of challenges for applying these methods in spatial knowledge extraction process. Like early years of GIS development, collecting data on a repository is still applied for knowledge extraction. In adaption, interoperability and usability are two important issues that have been neglected in traditional knowledge extraction methods. To respond the requirements from knowledge perspective a set of Web service-based geographical knowledge extraction is necessary. The services required a technological infrastructure or framework to extract useful knowledge from the different and geographically distributed data sources. In order to overcome this challenges, a general Spatial Knowledge Infrastructure (SKI) proposed to facilitate knowledge extraction in distributed scenarios. The SKI is a widespread interoperable framework through which geographical/spatial knowledge is created, organized, shared, managed and used in many domains. It creates a mechanism to make the necessary processes of geographic knowledge with the highest efficiency and usability. SKI's primary goal is to combine Spatial Data Infrastructure (SDI), Spatial Web Services (SWS) and Spatial Data Mining (SDM) concepts to facilitate geographical knowledge extraction as a collection of services. With a combination of SDI, SWS and SDM concepts this study presents the nature of SKI, including definition, levels, components, and architecture.
Keywords: Spatial Knowledge Infrastructure (SKI), Spatial Data Mining (SDM), Spatial Data Infrastructure (SDI), Spatial Web Services (SWS).

1 Introduction

Geographical Information Systems (GIS) has been evolving over the past few decades (Steiniger and Hunter, 2013; Longley *et al.*, 2015). Due to the development of new technologies including remote sensing, Internet of Things (IOT), smartphones and other geospatial concepts/cultures such as Spatial Data Infrastructure (SDI), Volunteered Geographic Information (VGI) and Public Participation GIS (PPGIS), the volume, number, and domain of geographical data have been growing increasingly (Miller and Han, 2009; Dahal and Chow, 2015; Li, Wang and Li, 2016). With the modern data collection technologies, geographic data accumulate at an alarming speed and now we can obtain much more diverse, dynamic, and detailed spatial data than ever possible before (Longley *et al.*, 2015). At the same time, with the continuing efforts by scientific projects, government agencies, and private sectors, voluminous geographic data have been collected, and continue to be archived. Accordingly, geographical sciences have moved from a data-poor age to a data-rich age (Miller and Han, 2009) that in this context an essential requirement is a method for discovering new knowledge in this voluminous geographic data (Longley *et al.*, 2015).

Although the vast availability of geographic data provides more opportunity for acquisition of new knowledge and a better understanding of complex geographical, environmental and social phenomena (Mansourian *et al.*, 2006; Toomanian, 2012), challenges related to applicability and usability of the data are remained and need to be addressed. Nowadays, there is a vital need for effective and efficient methods to extract knowledge from the voluminous geographic data (Mennis and Guo, 2009; Li, Wang and Li, 2016). In this regards, a number of Spatial Data Mining (SDM) methods and algorithms have been developed by the researcher during the recent years. Despite remarkable progress in SDM methods and algorithms, there are two main challenges involve the use of SDM in Geographic Knowledge Discovery (GKD) process. First, existing methods focus on a centralized perspective and did not consider distributed nature of geographic data. This means that like the early years of GIS development, collecting data on a repository is still applied for knowledge extraction. This limited perspective because of its limitations including distributed nature of geographical data, data sharing problems, reliability issues, and technological constrains is not well-suited for the GKD purposes. Second, interoperability and usability are two important issues that have been neglected in traditional SDM methods.The continuing advances in GIS technologies (i.e. geospatial services) and newly emerged data types such as Keyhole Markup Language (KML), Geography Markup Language (GML), Extensible Markup Language (XML), GeoRSS, GeoJSON, or other XML-encoded geo-formats promote interoperability concept in GIS technologies. However, the use of geospatial Web services to extract useful knowledge have not considered in previous SDM researches. Utilization of SDM methods based on geospatial Web services led to a usable and interoperable knowledge which can play an important role in GIS communities.

Web service-based GKD required a technological infrastructure or framework to extract useful knowledge from the different and geographically distributed data sources. Increasingly widespread use of the Internet provide new usability paradigm along with open standards (Li, Dragićević and Veenendaal, 2011; Steiniger and Hunter, 2013; Yue, 2013). The Open Geospatial Consortium (OGC) has published a number of standards include Web Map Service (WMS), Web Feature Service(WFS), Web Coverage Service (WCS), Web Processing Service (WPS) and etc., to simplification of GIS functionality as a service (Toomanian, 2012). In the current situation, it can be said that GIS is quickly evolving from a GI systems to a GI services. these services are designed to support interoperable spatial enable function over a network (i.e. Internet or Intranet) (Rajabifard, Feeney and Williamson, 2002; Mansourian *et al.*, 2006; Toomanian, 2012), but current services (and combination of them) emphasis on data and do not consider the GKD purposes. To respond the GIS community requirements from the knowledge perspective a set of Web service-based GKD is necessary. In order to overcome the challenges a general Spatial Knowledge Infrastructure (SKI) proposed to facilitate geographically distributed knowledge extraction in distributed scenarios. The focus of the study is introducing such infrastructure from the software and technological infrastructure perspective. In the following sections, we explain SKI definition, concept, and components.

2 Spatial Knowledge Infrastructure[1]

Spatial knowledge is a vague term used in many disciplines and there are many definitions for this phrase (Golledge, 2002; Laurini, 2014). In general, it is defined as interesting high-level information or facts to solve geographic problems in various domains (Longley *et al.*, 2015). In other words, understanding four questions concerning geographic entities include "why", "how", "what" as well as "where" lead to a perception and understanding nature of geographical phenomena and creation of spatial knowledge or rules. In this regards, geometry and topology are two important features for determination of spatial relationships between objects (Mennis and Guo, 2009; Miller and Han, 2009). This kind of facts are extremely important for economic, social and environmental problems and are stated as the main elements of sustainable development (Rajabifard, Feeney and Williamson, 2002; Mansourian *et al.*, 2006; Toomanian, 2012).

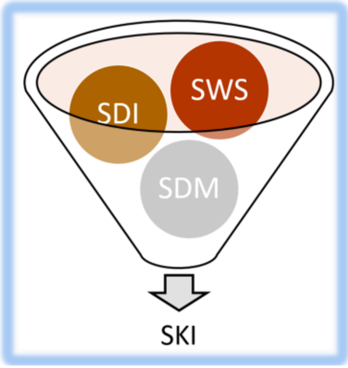
To extract useful knowledge from distributed spatial data sources, a knowledge infrastructure is required (Kargupta, 2009; Talia and Trunfio, 2013). From a technological perspective, knowledge infrastructure is an integrated set of technological components (typically hard components include computer and networking hardware and facilities, and soft components also include various software, services, procedures, protocols or standards) that are the foundation of a knowledge service. This means that without such infrastructure it is not possible to facilitate knowledge management process (e.g., create, share, access, and manage knowledge). Besides, the development of new communication technologies (i.e., Web services, distributed framework, and IOT) and GIS (i.e., ubiquitous-GIS), the nature of spatial/ geographic[2] knowledge has evolved over the past few decades (Golledge, 2002). Todays, with the advent of knowledge society concepts, seamless access to the right spatial knowledge at the right time, in the right format (geo-products and geo-services), at the right price due to a SKI in a distributed environment, and offer some solutions and becoming more vital. In this regards, spatial knowledge can be conceptualized as a collection of services, this means that the concept of spatial knowledge has shifted from its traditional mode (e.g., spatial analysis process) to a new generation of concepts named Spatial Knowledge Services (SKS).

Although, SDI is well-suited for distributed data-driven processes and services (Rajabifard, Feeney and Williamson, 2002; Williamson, Rajabifard and Feeney, 2003; Mansourian *et al.*, 2006; Toomanian, 2012) include Extraction, Transformation, Load (ETL) and sharing spatial data, it is not adapted for knowledge extraction in interoperable environment. This concept can be enlarged to an SKI based on the above vision. The SKI is a widespread interoperable framework through which spatial knowledge is created, organized, shared, managed and used in many domains. It creates a mechanism to make the necessary processes of geographic knowledge with the highest efficiency and usability. SKI's primary goal is to combine SDI, SWS and SDM concepts to facilitate GKD as a collection of services (Figure.1).

[1] SKI refers to a broad concept, and this study focuses on the technological aspect of it.

[2] There are some differences between these terms, spatial knowledge related any spatial insight in 2 or 3 dimensional space, but geographical knowledge related to insights about earth or near surface of the earth.

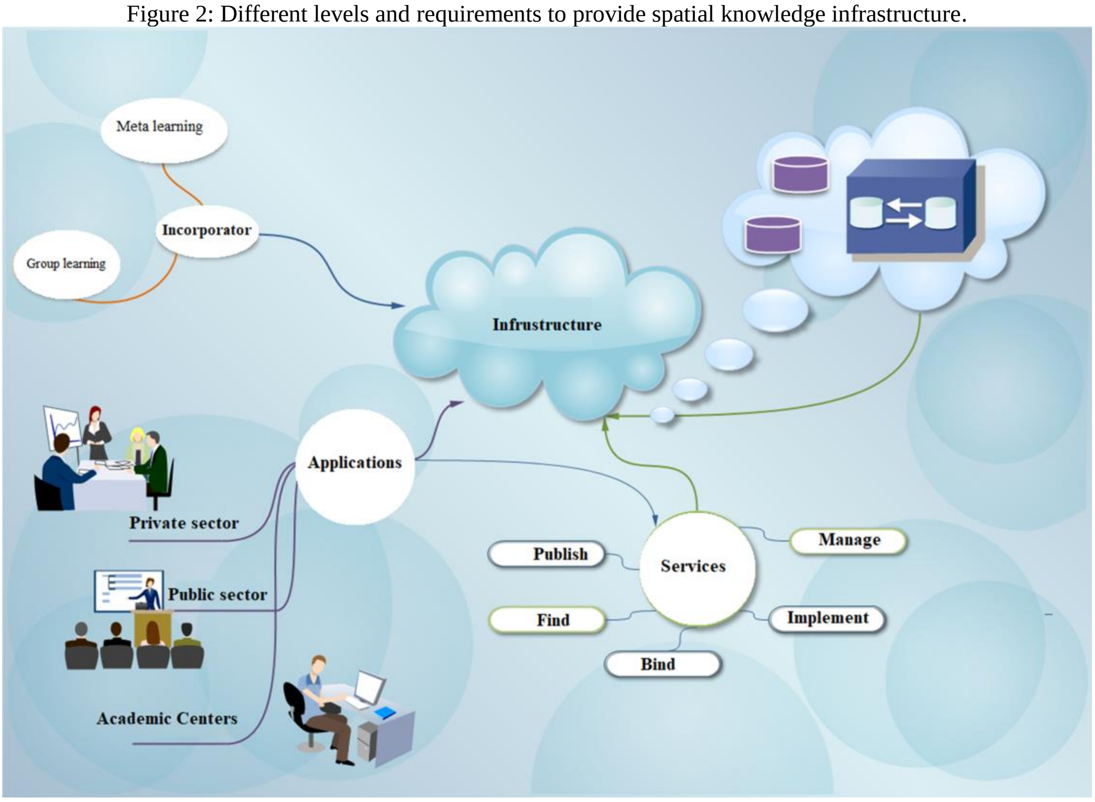
Figure 1: Combination of the SDI, SWS and SDM concepts in order to create an SKI.



3 Spatial Knowledge Infrastructure

In the previous section, a brief definition of the SKI was presented. In this section, more details of the SKI components and relationships among them are described. The SKI includes a set of components, such as software, hardware, data, procedures, standards, and services that interact with each other to create, explore and sharing useful knowledge from geographic data. Considering the fact that GIS and other geo-informatics technologies are already moving towards a network-based systems (Kargupta, 2009; Li, Dragićević and Veenendaal, 2011; Yue, 2013), development of Service Oriented Approaches (SOA), and new knowledge-based infrastructure have have been emerging as an important strategy for enhancing GKD process. Only through such services can one expect that GIS change from a data-centric, centralized, and static paradigm to a service-based, ubiquity-based, and knowledge-based paradigm. Through the SOA approach, spatial information communities are able to define integrated services that facilitate knowledge extraction from spatial data. Such an approach can be found in a multitier architecture and at four general levels include level #1 Infrastructure, level #2 Service, level #3 Incorporator and Level #4 Application (Figure 2).

Figure 2: Different levels and requirements to provide spatial knowledge infrastructure.



3.1 Infrastructure

Infrastructure is a set of structural and interconnected technological factors that allow for the creation of other levels. Without such a structure, other levels cannot be operationalized easily or effortlessly (Talia and Trunfio, 2013). Here, the knowledge infrastructure is needed to achieve an enterprise SKI. In a knowledge infrastructure, computing hardware, data sources, networks, digital sensors, and online observational tools are merged with a software suite and middleware tools and services. This component can be provided in the Grid or Cloud environment (Kargupta, 2009). These capabilities include security services (e.g., authentication, authorization, encryption), data management services (access to data, ETL process or file transfer and etc.), and executive management services (e.g., resource allocation, process creation) and knowledge services (publish knowledge services).

3.2 Services

In this level, knowledge discovery interfaces related to SKS, specified to create, discover and share. These interfaces are a set of operations that characterize the behavior of SKS. Such services provide a mechanism for describing, finding, binding and retrieving knowledge from geographic data based on SDM algorithms or other GKD methods. Consider to reusability and interoperability are the most important features that should be reflected in the implementation phase (Nalepa and Furmańska, 2009; Li, Dragićević and Veenendaal, 2011). In this regards, several standards maintenance this deployment include Web Services Description Language (WSDL), Security Assertion Markup Language (SAML), Universal Description, Discovery, and Integration (UDDI), and Simple Object Access Protocol (SOAP). These standards providing the definition of Web services and their advertisement to the potential user community, binding for invocation purposes, and reuse (Sheth, Benslimane and Dustdar, 2008). At the same time, OGC has also developed various standards for geospatial web services (Li, Dragićević and Veenendaal, 2011; Toomanian, 2012). Due to development of this standard or protocols reusable and interoperable service output product can be used in many domains and applications.

3.3 Incorporator

In this level, more complex and high-level spatial knowledge created through a combination of the services. The most important features of this level are group and meta learning (Kargupta, 2009). For example, in SDM process one may use province X data for a model learning and Y province data for a model-testing scenarios. Also, the combination of knowledge extracted at different levels can be imagined.
Semantics and Ontology based Web services can be used on the incorporator level for organizing knowledge exchange between users (Pradhan and Varde, 2016). The Semantic Web offers a general interoperable context in which evidence is specified a well-defined meaning such that facts and applications can be handled by machines for more useful finding, binding, integration and reuse through many applications (Yue, 2013; Jelokhani-Niaraki, Sadeghi-Niaraki and Choi, 2018).

3.4 Applications

Since the SKI products presented as a set of services, interaction between distributed and heterogeneous applications are possible. Therefore, the products can be used in a wide range of applications include, public/ private organizations, software products and, even geoportals.

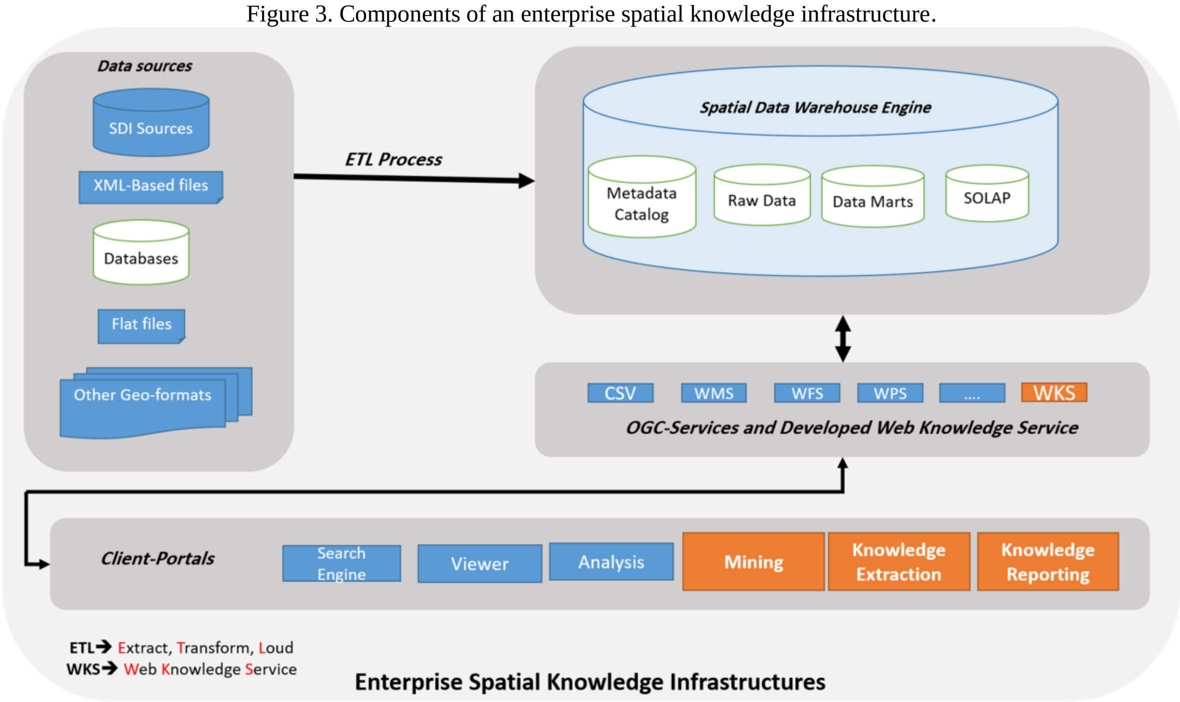
Using service-oriented capabilities enables a user to extract knowledge from data and then share the knowledge gained in the previous step. It is also possible through a combination of different knowledge create a meta-knowledge. It should be noted that achieving such a goal associated with many complications because knowledge is much related to the subject and its domain. Although the initial cost of creating such an infrastructure and services are costly or out of organization budget, through sale of the services stakeholders are able to earn a steady income.

4 Components and architecture of SKI

The main SKI components include data, ETL, spatial warehouse engine, Spatial On-line Analytical Processing (SOLAP) server and a collection of web knowledge services (see Figure 3). The primary component in the SKI architecture is data. Here, the nature of data is distributed and heterogeneous (different formats and a variety of different sources). Although, the main objective of SKI is not collecting data in a single repository, while using spatial services to extract knowledge from distributed data sets, an online storage repository is required. At this stage, the combination data from heterogeneous resources is the main challenge (Kargupta, 2009; Miller and Han, 2009). Therefore, a component named ETL is required for converting, integrating, rendering, and finally storing heterogeneous data in a data warehouse (component #3). The availability and accessibility of data in a homogenous data warehouse allows to implement advance query (i.e., cubes) or SDM algorithms as a service (Miller and Han, 2009). Using an SOLAP server, it is possible to make a multi-dimensional query from geographic data (Mennis and Guo, 2009). In fact, this component provides the consistent integration of spatial objects into the SOLAP data cube structure. The SOLAP server connects to data repository and performs SOLAP operations using Multidimensional Expressions (MDX) queries. The fifth component of the SKI is the web knowledge services. These services are also capable of performing SDM algorithms and provide facilities for summarizing data. The most important features of these web services are:

- *Dynamically:* discovering, applying and sharing SDM algorithms are utilized dynamically in a service-based framework.
- *Scalability:* The Web services provide the ability to combine new data sources, new algorithms, and new tools in any platform.
- *Reducing complexity:* Compared to traditional GKD methods, the user focuses on what he/she wants, instead of focusing on how to do it and its details. As a result, the proposed framework greatly reduces the complexity of the exploration of knowledge from heterogeneous and distributed geographic data.
- *Extendibility:* to extracting knowledge from geographic data various services can be defined for different algorithms and methods.

Figure 3. Components of an enterprise spatial knowledge infrastructure.



5 Conclusion

From a technological perspective, spatial knowledge can be conceptualized as a collection of services. To extract, exchange and integrate useful knowledge from distributed spatial data sources a knowledge infrastructure is required. With a combination of SDI, SWS and SDM concepts, this study presents the nature of SKI including definition, levels, components and, architecture. It produces a mechanism to make the essential procedures of geographic knowledge with the maximum efficiency. Such framework, presented at four general levels including infrastructure, service, composition and, application. Moreover, the components of an enterprise SKI was proposed. Although, SKI is a broad concept (e.g., technological, economic, socio-cultural, organizational and governmental) this study, focused on the technological context of SKI, accordingly, other dimensions can be studied in a particular research.

6 References

Dahal, K. R. and Chow, T. E. (2015) ‘Characterization of neighborhood sensitivity of an irregular cellular automata model of urban growth’, *International Journal of Geographical Information Science*. Taylor & Francis, 29(3), pp. 475–497.

Golledge, R. G. (2002) ‘The Nature of Geographic Knowledge’, *Annals of the Association of American Geographers*. Blackwell Publishing Inc, 92(1), pp. 1–14. doi: 10.1111/1467-8306.00276.

Jelokhani-Niaraki, M., Sadeghi-Niaraki, A. and Choi, S.-M. (2018) ‘Semantic interoperability of GIS and MCDA tools for environmental assessment and decision making’, *Environmental Modelling & Software*, 100(Supplement C), pp. 104–122. doi: https://doi.org/10.1016/j.envsoft.2017.11.011.

Kargupta, H. (2009) *Next generation of data mining*. CRC Press.

Laurini, R. (2014) ‘A conceptual framework for geographic knowledge engineering’, *Journal of Visual Languages & Computing*. Academic Press, 25(1), pp. 2–19. doi: 10.1016/J.JVLC.2013.10.004.

Li, D., Wang, S. and Li, D. (2016) *Spatial data mining : theory and application*. Available at: http://www.springer.com/us/book/9783662485361 (Accessed: 28 August 2017).

Li, S., Dragićević, S. and Veenendaal, B. (2011) *Advances in web-based GIS, mapping services and applications*. CRC Press/Balkema.

Longley, P. A. *et al.* (2015) *Geographic information science and systems*. 4th edn. Wiley. Available at: https://www.amazon.com/Geographic-Information-Science-Systems-Longley/dp/1118676955.

Mansourian, A. *et al.* (2006) 'using SDI and web-based system to facitilate disaster management', *Computers & Geosciences*, 32(3), pp. 303–315. doi: http://dx.doi.org/10.1016/j.cageo.2005.06.017.

Mennis, J. and Guo, D. (2009) 'Spatial data mining and geographic knowledge discovery-An introduction', *Computers, Environment and Urban Systems*, 33(6), pp. 403–408. doi: 10.1016/j.compenvurbsys.2009.11.001.

Miller, H. J. and Han, J. (2009) *Geographic data mining and knowledge discovery*. CRC Press.

Pradhan, A. M. and Varde, A. S. (2016) 'Ontology based meta knowledge extraction with Semantic Web tools for ubiquitous computing', in *Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), IEEE Annual*, pp. 1–6.

Rajabifard, A., Feeney, M.-E. F. and Williamson, I. P. (2002) 'Future directions for SDI development', *International Journal of Applied Earth Observation and Geoinformation*. Elsevier, 4(1), pp. 11–22.

Sheth, A., Benslimane, D. and Dustdar, S. (2008) 'Services Mashups: The New Generation of Web Applications', *IEEE Internet Computing*, 12, pp. 13–15. doi: 10.1109/MIC.2008.110.

Steiniger, S. and Hunter, A. (2013) 'The 2012 free and open source GIS software map–A guide to facilitate research, development, and adoption', *Computers, Environment and Urban Systems*. Available at: http://www.sciencedirect.com/science/article/pii/S0198971512000890 (Accessed: 3 May 2017).

Talia, D. and Trunfio, P. (2013) *Service-oriented distributed knowledge discovery*. CRC Press.

Toomanian, A. (2012) *Methods to Improve and Evaluate Spatial Data Infrastructures*. Lund University, Sweden. Available at: http://www.lunduniversity.lu.se/lup/publication/40d993fc-98ca-4f12-becd-a0e8d25c9048.

Williamson, I. P., Rajabifard, A. and Feeney, M.-E. F. (2003) *Developing spatial data infrastructures : from concept to reality*. Taylor & Francis.

Yue, P. (2013) *Semantic web-based intelligent geospatial web services, springer*. Available at: http://link.springer.com/content/pdf/10.1007/978-1-4614-6809-7.pdf.