Geospatial Data Sharing Based on Geospatial Semantic Web Technologies

C. Zhang
W. Li
T. Zhao

Geospatial data sharing is a concern in geospatial science because of the heterogeneity of existing geographical information systems. This study aims to examine the use of Geospatial Semantic Web technologies such as ontology, web services and the service-oriented architecture for enabling disparate heterogeneous legacy GIS to share and integrate information in a cost effective way to reduce spatial data duplication. A framework based on the Geospatial Semantic Web technologies is proposed in this study. Experimental results from an implemented prototype show that the proposed framework allows searching and accessing geospatial data and services at the semantic level based on their content instead of keywords in the metadata.

Key words: Internet GIS, web services, data sharing, ontology, geospatial semantic web

C. Zhang
W. Li
Department of Geography
Kent State University,
Kent, Ohio 44242, USA
czhang2@kent.edu
weidong6616@yahoo.com
T. Zhao
Department of Computer Science,
University of Wisconsin, Milwaukee, USA
tzhao@uwm.edu

INTRODUCTION

Many legacy Geographic Information Systems (GIS) have been developed over different periods, for different purposes, by different organizations and with different structures and vocabularies based on different GIS software, such as ESRI ArcInfo and ArcView, Smallworld GIS, Intergraph GeoMedia, MapInfo professional, and Clark Lab's Idrisi. The legacy GIS built on different GIS software have their own proprietary system designs, data models and database storage structures. Thus, geographical databases based on these systems cannot communicate without data conversion. However, data conversion is costly and time consuming and may lead to the compatibility problems for many time-critical applications such as emergency response, location-based service, and real time traffic management, which need real-time access to diverse data to make quick decisions and take instantaneous actions (Zhang and Li, 2005). Although the development of the World Wide Web (WWW) and many Internet GIS provide proprietary ways to allow users to quickly access, display and query spatial data over the web, these Internet GIS also have the limitations of proprietary software designs, data models and database storage structures. Mapping and geoprocessing resources distributed over the web by these Internet GIS cannot be shared and interoperated in real time. Data sharing facilitated by the advances of network technologies is hampered by the incompatibility of the variety of data models and formats used at different sites (Choicki, 1999).

Data with different formats are often remotely accessible only through simple protocols (e.g., FTP) that do not allow queries or filtering, which are difficult to integrate with different applications. Further, the heterogeneity of GIS and the difficulty to share geospatial information cause the spatial data duplication problem. Redundant efforts are commonplace in the legacy GIS. Although the geospatial community realizes the costs of spatial data duplication and the heterogeneity of existing GIS and applications, which are seriously limited in sharing spatial data and thus inhibit the synchronization between GIS and businesses, they cannot simply discard these GIS and start new systems from scratch because it would be too time consuming and costly. A solution that builds on initial expenditures rather than starting a new initiative so as to allow agile integration of these GIS and sharing geospatial data in a cost effective way, is needed.

To facilitate the exchange and sharing of spatial data by building on initial expenditures, Spatial Data Infrastructures (SDI) have been developed in many countries in the past two decades (e.g. Crompvoets *et al.*, 2004; Masser, 2005; Rajabifard *et al.*, 2006). Recent SDI development is based on the open standards and Open Geospatial Consortium (OGC) web service technologies (GSDI, 2004; OGC, 2003a). The use of open standards and OGC web service technologies offer the potential to overcome the heterogeneous problems of legacy GIS and sharing geospatial data in a cost-effective way (Askew *et al.*, 2005; Peng and Zhang, 2004a; Peng and Zhang, 2004b; Zhang *et al.*, 2003; Zhang and Li, 2005; Tait, 2005). Significant progress has been made in terms of implementation of global, regional and local SDI based on the open standards and OGC web services (e.g. Williamson *et al.*, 2003; Crompvoets *et al.,* 2004; Tait, 2005; Mansourian *et al.*, 2006).

Although the fast development of SDI and OGC web service technologies has undoubtedly improved the sharing and synchronization of geospatial information across the diverse resources, literature of SDI shows that there are limitations of the current implementation: 1) the implemented SDI only emphasize technical data interoperability via web services and standard interfaces and cannot resolve semantic heterogeneity problems in spatial data sharing; and 2) with current implemented SDI it is only possible to search and access geospatial data and services by keywords in metadata and it is impossible to directly search and access geospatial data and services based on their content. Difference in semantics used in different data sources is one of the major problems in spatial data sharing and data interoperability (Bishr, 1998; Fabrikant and Buttenfield, 2001). One possible approach to overcome the problem of semantic heterogeneity is by means of ontology (Fonseca *et al.*, 2002; Pundt and Bishr, 2002; Smith and Mark, 1998). The concept of the Geospatial Semantic Web was suggested recently to address the vexing semantic challenges and achieve automation in service discovery and execution (Peng and Zhang, 2005). Geospatial Semantic Web can be seen as an extension of the current web where geospatial information is given well-defined meaning by the ontology, thus it allows users to have a semantic spatial query capability (Duke *et al.*, 2005). This study aims to examine the use of Geospatial Semantic Web, OGC web services and Service Oriented Architecture (SOA) for enabling disparate GIS to share and integrate geospatial information at the semantic level in a cost effective way. Thus the systems built on these technologies can search and access geospatial data and services by their content rather than just by keywords in metadata. A framework based on Geospatial Semantic Web, OGC web services and SOA technologies has been proposed in this study for integration of legacy GIS and geospatial data sharing. A prototype has been implemented to demonstrate the feasibility of searching and accessing geospatial data and services based on their content instead of keywords in the metadata. And it allows sharing geospatial data from heterogeneous databases at the semantic level over the web through ontologies and OGC web services. The experimental results of the implemented prototype demonstrate that ontologies and OGC web services can be a way to enable semantic interoperability and to reuse heterogeneous spatial data from disparate legacy GIS. Also, that SOA provides more flexible and loose coupling of GIS resources than traditional system architectures and permits an easy way of data integration.

A Framework of Geospatial Data Sharing Systems with Geospatial semantic web technologies

A framework of geospatial data sharing systems with Geospatial Semantic Web Technologies for heterogeneous legacy GIS is proposed as shown in Figure 1. For instant remote data access and exchange, the ontology-based web services are used to access and manipulate geospatial data through the web from heterogeneous databases. This approach ensures basic conditions for interoperability by using a standard exchange mechanism and Geospatial Semantic Web Technologies between diverse spatial data sources connected over the web. The main advantages of this approach are 1) to guarantee data interoperability at semantic level for geospatial data sharing over the web; and 2) to search and access geospatial web services by content rather than just by keywords in metadata. Applications and organizations using this approach can deploy spatial data with different semantics over the web in real time so that information from diverse sources with incompatible data formats and semantics can work together transparently across the web. The following subsections discuss some technologies used in the proposed framework.

A Semantic Service-Oriented Architecture

The framework is based on Service-Oriented Architecture and is essentially a collection of OGC web services, which communicate with each other by simple data passing or coordinating some activities. It is composed of four elements: service provider, service broker, service client, and ontology server. Figure 2 illustrates the four components of the proposed framework in the logic universe. The service provider supplies geospatial data; the service client allows users to search and integrate data from providers; and the service broker provides a registry for available services. The ontology server ensures

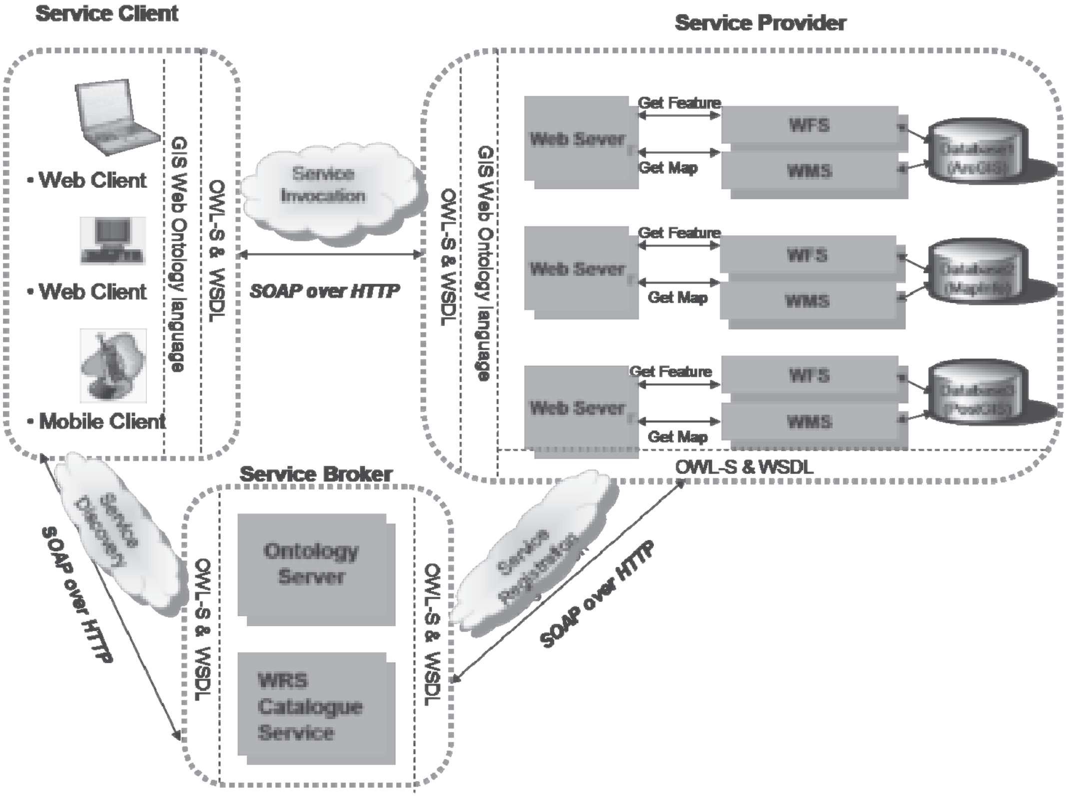


Figure 1. *A framework of geospatial data sharing systems with Geospatial Semantic Web technologies.*

the semantic interoperability of the ontologies of service clients and providers.

The service provider uses OGC data services such as Web Feature Service (WFS) (OGC, 2005) and Web Map Service (WMS) (OGC, 2004a) to publish heterogeneous geospatial data connected to a legacy GIS. The OGC data services provide a basis to share spatial data with different data models and from different sources without data conversation. The service broker uses the ontology-based OGC Web Register Service (WRS) and Catalogue Service (CS) (OGC, 2004b) to register and manage the data services and allow users to search for these data services. Service clients search contents of catalogue services to find the datasets and services of interest, and they also can combine the found data services through catalogue web services. Unlike traditional web services, our service clients and providers must maintain local ontology at both the client side and the provider side to ensure the semantic interoperability. Local ontology at the service client refers to semantics used by the client user or client applications, while local ontology at the service provider refers to semantics used by the data providers. These local ontologies may address geospatial relations such as topological relations (e.g. connectivity, adjacency, and intersection among geospatial objects), cardinal directions (e.g. east, west, northeast), and proximity relations (e.g. the geographical distances among objects). The client ontology must be able to communicate with the provider ontology. Moreover, a service client may need to access multiple providers to complete a task. Thus, it is necessary to create mappings of equivalent or related classes and properties in the local ontologies. The ontology server is used to realize this function and it keeps a taxonomy of geospatial terminologies and maintains consistency for different local ontologies. The service broker uses the ontology server to map the standard OGC catalogue services to ontology-based catalogue services.

The web services are connected via OWL-S (Semantic Markup for Web Services) and Web Service Description Language (WSDL) among the service provider, the service broker and the service client. OWL-S (http://www.daml.org/services/owl-s/1.0/owl-s.html) is an ontology of services built on the Web Ontology Language (OWL) to facilitate automatic web service discovery, invocation and composition. OWL-S uses a process model to describe services. The process model contains a number of atomic processes that can be invoked individually or combined together. The data used by the processes are based on description logic types. Since the standard OGC web services are based on WSDL and geospatial web services use WMS/WFS as extensions of WSDL, rules have been implemented to translate the OWL-S processes from/to WSDL operations and the input/output values of the processes from/to WSDL messages. The Simple Object Access Protocol (SOAP) binding over HTTP is employed for communication between web services via the Internet. SOAP essentially provides the envelope for sending web services messages.

In general, the semantic SOA in the proposed framework ensures data interoperability through semantic web services, which offer basic conditions for interoperability by using a standard exchange mechanism between diverse spatial data sources connected over the web. The semantic web services provide the interoperable capability of cross-platform and cross-language functionality in the distributed

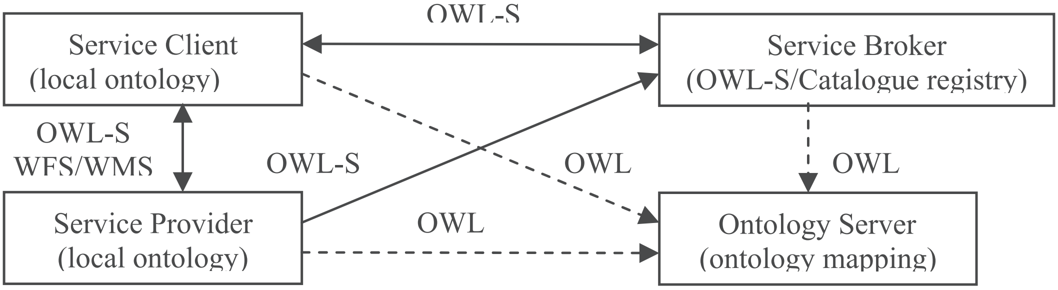


Figure 2. *Service components of Geospatial Semantic Web in the logic universe.*

Internet environment. The semantic web services are accessed via ubiquitous web protocols (e.g. HTTP) using universally accepted data formats (e.g. XML), and they represent functionality that can be easily reused without knowing how the service is implemented.

Semantic Descriptions of Geographic Information using Ontologies

Several XML-based mark-up languages are available for the description of ontologies including RDF (Resource Description Framework) Schema (W3C, 2004) and Web Ontology Language (OWL) (Antoniou and Van Harmelen, 2003). The OWL is adopted in the framework to create the web-based ontologies. The OWL draws upon the formal theory of Description Logic (DL), which has roots in first-order predicate logic and provides highly expressive concept-forming constructs (Baader *et al.*, 2003).

The web-based ontologies are stored in an ontology server. Both service clients and providers can register their application ontologies to the ontology server. The ontology server ensures the interoperability of the ontologies of service clients and providers. As the ontologies may be created by different communities, thus heterogeneity problems may arise when integrating the information from two or more ontologies. The G-Match algorithm proposed by Hess *et al.* (2006) is adopted in the framework to integrate geographic ontologies in the ontology server. The algorithm measures the similarities of the concepts of two different geographic ontologies by considering their names, attributes, taxonomies, and conventional as well as topological relationships. Except for the name comparison, G-Match considers both the commonalities and the differences for measuring the similarity between two concepts. In the G-Match algorithm, each concept c_i from an ontology O is compared against all concepts c_j from ontology OO. The G-Match has three main phases of similarity measure in its execution: first, the concept names (SimName(ci),cj) and attributes (SimAt(ci),cj) are compared. Then, using the results from the name similarity measure, the taxonomy (SimTx(ci),cj) relationships (SimRel(ci),cj) and topology relationships (SimTop(ci),cj) are evaluated. Finally, the overall similarity measure is estimated.

OWL-S is used to create semantically enriched web feature service and web map service descriptions in this framework. OWL-S is an upper-ontology based on OWL that can model the characteristics of web feature services and web map services. OWL-S provides three modeling constructs at the top level, i.e., the service profile (what the service does), the service grounding (how the service can be accessed) and the service model (how to use the service in terms of semantic content). Through the three modeling constructs OWL-S overcomes problems caused by semantic heterogeneity in OGC data services and supports automatic discovery, composition, invocation and orchestration of these data services.

The Ontology-based Catalogue Service for Service Discovery

Finding suitable spatial information in the open and distributed web environment is a crucial task. The current practice of the implemented clearinghouse and geoportal is to search user interested geospatial data and services based on keywords in metadata. While it is useful to obtain the geospatial data by keywords-based search for metadata, metadata still has semantic heterogeneity problems. Different metadata creators may use different names for the same feature. In addition, metadata contains only limited information to allow users to search. To overcome limitations of searching metadata and allow users directly search for geospatial data and services, the proposed framework adopts the ontology-based service broker (or catalogue service). Unlike the traditional service broker, such as the OGC catalogue service that enables discovery and retrieval of metadata, the ontology-based service broker (or catalogue service) in the framework provides searchable repositories of service descriptions at the semantic level and allows users to directly search and access geospatial information based on semantic content. Thus instead of providing resource-based metadata that describes services and data sets, the ontology-based service broker can allow users to directly locate, access and make use of geospatial services and data sets in an open, distributed system.

The ontology-based catalogue service in the framework has two main functions -- discovery and publication. *Discovery* means that the service requester seeks to find resources of interest through simple browsing or by sophisticated query-driven discovery that specifies simple or advanced search criteria. The ontology-based catalogue performs the search and returns a result set which contains all registry geospatial services and data sets that satisfy the search criteria. The requester may then choose to retrieve representations of result set items according to some specified schema and element set. The ontology-based catalogue service in the framework allows web services to *publish* resources in two ways: harvest operation and transaction operation. The *harvest operation* is the basic publication activity and in this operation the catalogue itself attempts to harvest a resource from a specified network location, thereby realizing a *pull* model for publishing registry geospatial services and data sets. If the catalogue successfully retrieves the resource and can handle it, then one or more corresponding registry services and data sets are created or updated. The *transaction* operation provides support for modifying catalogue content through a public interface that allows a *push* style of publication. A user may insert, update or delete catalogue entries according to criteria specified in the request message. This operation is subject to some kind of access control such that only authorized users may perform such actions.

The main advantage of the ontology-based catalogue service is that it overcomes semantic heterogeneity caused by synonyms and homonyms during keyword-based searching in standard OGC catalogues, which has been acknowledged by many studies (e.g. Fonseca *et al.*, 2002; Cruz, *et al.*, 2002; Klien *et al.*, 2005; Lutz and Klien, 2006). The ontology-based catalogue service allows semantic content-based searching by combining several technologies including standard markup languages, ontologies and knowledge bases via semantic reasoning services.

The DL is used as the formalism for knowledge representation in the ontology-based catalogue services. It supports intersection, union, full existential quantification, value restriction, negation and number restrictions to inductively construct complex concept descriptions out of primitive concepts and roles. The reasoning services provided by DL allow one to perform a knowledge-based search in the proposed framework. The common shared ontologies in the ontology server provide the necessary requirements towards logic-based matchmaking search. Through terminological reasoning, the ontology-based catalogue services can help to find appropriate matched geospatial web services or data sets. The ontology-based catalogue services in combination with terminological reasoning ensure automatic discovery at the semantic level of geospatial web services and data sets.

The Ontology-based Geospatial Information Retrieval

Recently two approaches were proposed for geospatial information retrieval—subsumption reasoning and similarity measurement (Janowicz, 2006). The similarity measurement approach proposed by Janowicz (2006) is adopted in the proposed framework to retrieve geospatial information at semantic level. The main idea behind similarity measurement retrieval is to compute the degree of overlap between search and compared-to concept to simplify phrasing an adequate search concept and the results are ordered by their degree of similarity to the searched concept. The advantage of similarity measurement retrieval is that it can deliver concepts that are close to, but not identical with the user's intended search concept. It supports a basic notion of context and conceptual neighborhood models, which are necessary to handle spatial and temporal relations.

In the similarity measurement retrieval, similarity between concepts in normal form is measured by comparing their description logics for overlap, where a high level of overlap indicates high similarity and *vice versa*. The complex concepts are specified out of primitive concepts and roles using given language constructors. The overall similarity between concepts is the normalized and weighted sum of the single similarities calculated for all parts of the concept descriptions. For example, to compute overall similarity (sim_u) between two concepts C and D, the similarity between the disjunctions C_1 Y...YC_n and D_1 Y...YD_m has to be measured using the following equation:

$$sim_u(C,D) = \sum_{(C_i,D_j)\in SI} \omega_{ij} * sim_i(C_i,D_j) \tag{1}$$

where (C_i, D_j) represents a tuple of simple concepts, SI represents the set of tuples (C_i, D_j), ω represents weights according to the probability of tuples (C_i, D_j). Note that the sum of ω all is always 1. Depending on the application area, ω can be computed out of the set cardinality (based on the relations between individuals and their concepts) of all involved concepts on disjunction level, using probability assumptions (based on both the relations between individuals and concepts, and the relations between concepts), or from the structure of the examined ontology (based on the relations between concepts).

A Prototype Implementation of the Framework

The framework discussed above covers several components, such as using OGC WFS and WMS services to publish the heterogeneous spatial data connected to legacy GIS, using ontologies to describe the geographic information, using ontology-based WRS and CS to register and discover the published WFS and WMS services, using OWL-S and WSDL as a service interface to connect service providers, service brokers and service clients, using SOAP over HTTP for communication between web services over the web, using a similarity measurement approach to retrieve the ontology-based geographic information and, finally, users share and integrate the heterogeneous spatial data by binding to WFS or WMS services. A prototype that contains these components was implemented based on the proposed framework shown in Figure 1. The main goals of the prototype include: (1) to search and access heterogeneous spatial data based on semantic content instead of keywords in metadata; (2) to integrate heterogeneous spatial data via OGC data services through ontology-based web register services and catalogue services; and (3) to illustrate the agility of SOA by flexible loose coupling of legacy GISystems. Figure 3 illustrates the architecture of the implemented prototype. The prototype's architecture consists of:

(1) Service providers:

- ESRI ArcGIS and PostGIS, which provide different format spatial data;
- Geoserver (http://geoserver.sourceforge.net/html/index.php), an open-source software which enables full implementation of the OGC WFS and WMS specifications and serves ShapeFile and PostGIS data using WFS and WMS;
- Java 2 Platform, Enterprise Edition (J2EE), the supporting environment for GeoServer;
- Apache HTTP server, which serves as a web server for WFS and WMS;
- Tomcat, a java servlet container, which provides web developers with a simple consistent mechanism for extending the functionality of a web server and for accessing web application GeoServer.

(2) Service brokers:

- An ontology server (http://zhangweb.geog.kent.edu/ontology/Transitontology.owl), which maintains consistency for different local ontologies and is developed using Protégé software (http://protege.stanford.edu/).
- An ontology-based catalog developed by ourselves based on ebXML and Java API for XML registries (JAXR).
- Oracle XMLDB, which provides capabilities for the storage and management of XML data for ontology-based catalog.
- Apache HTTP Server, which serves as a web server for web registration and catalogue services.
- Tomcat JSP and Servlet, which allows for accessing the ontology-based catalog.
- Java Virtual Machine, the supporting environment for the ontology-based catalog.

(3) Service clients:

- A Java-based WFS Client developed by us (http://jiangxi.cs.uwm.edu/wfs/), which provides an interface for query and access to ontology-based web services such as WFS and WMS. The Java-based WFS Client software is a web-based client software and allows users to query and access geospatial data at any time over the web from data providers.

The data used come from the Waukesha Transit Trip Planning Project, an online bus trip-planning website for the City of Waukesha, Wisconsin, USA. Two WFS/WMS servers are built in the implemented prototype to publish two different format geospatial data—ESRI Shapefiles and PostGIS database over the web from two separate servers. The data of bus routes, streets and landmarks/facilities (original format is ESRI Shapefiles) are located in the University of Wisconsin-Milwaukee computer (http://qingdao.cs.uwm.edu:8090/geoserver/wfs?service=WFS&request=GetCapabilities). The data of bus stops (original format is PostGIS database) are located in the Kent State University computer (http://zhangweb.geog.kent.edu:8080/geoserver/wfs?service=WFS&request=GetCapabilities). Note that these two different data formats are chosen just for test purposes.

In the implemented prototype, the general steps involved in providing and consuming a web service include:• A service provider creates WFS and WMS services, which are tied to the databases of legacy GIS ArcGIS and PostGIS. The description of WFS and WMS services uses WSDL. Then the WSDL is translated to the OWL-S processes. The WFS and WMS services are registered and published using OWL-S to a directory of ontology-based catalog services using SOAP messages over HTTP.

- A service consumer issues one or more queries through the Java-based WFS client to the directory of CS to locate a service. Initially, the client will translate user requests into OWL-S processes. Then it will consult the catalog services to find the appropriate providers, and use the ontology server to generate correct WFS/WMS messages.
- Part of the OWL-S for WFS and WMS services provided by the service provider is passed to the service consumer using SOAP messages

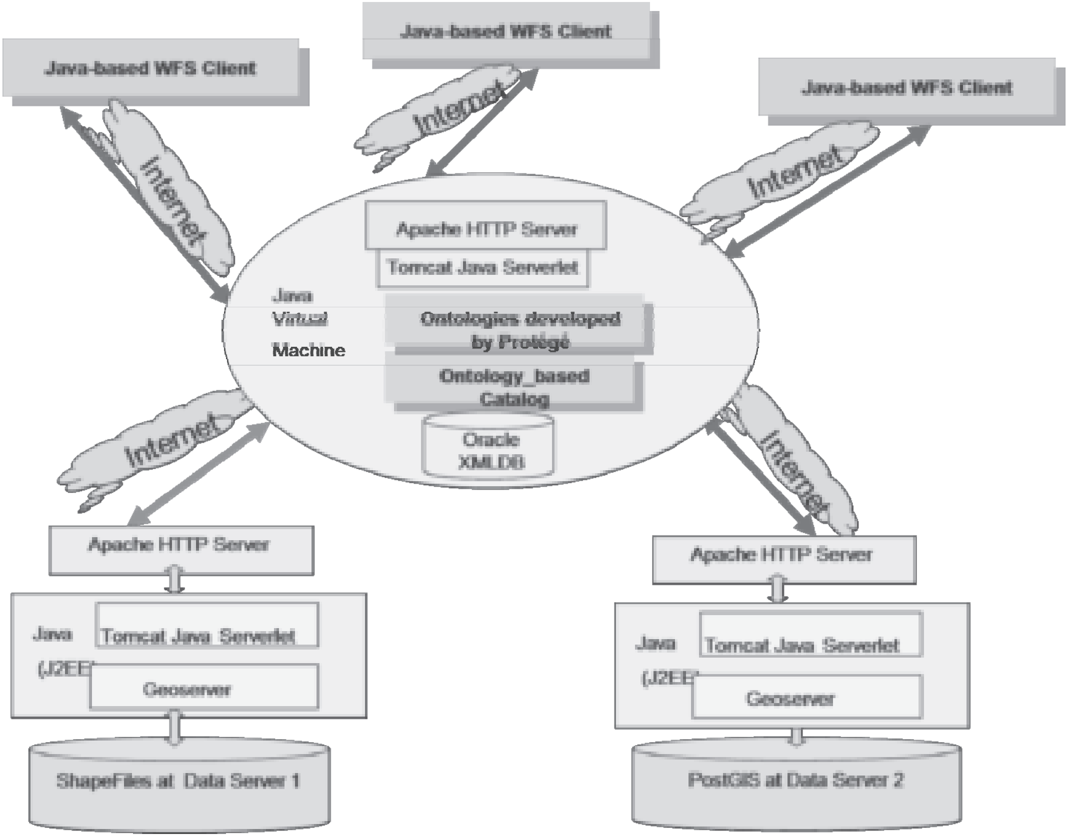


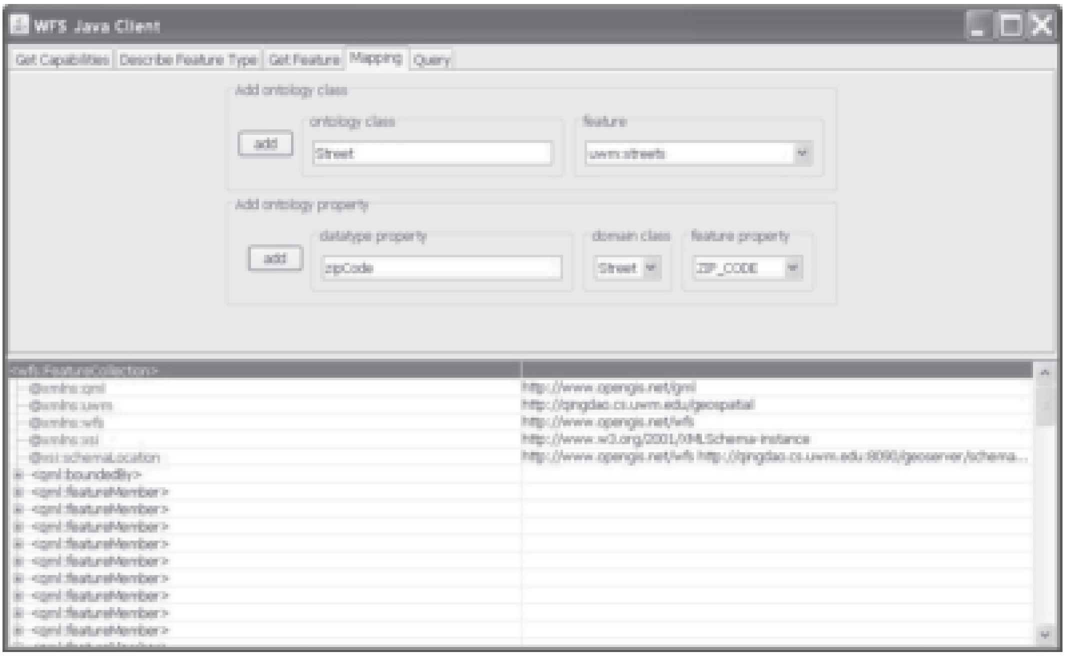
Figure 3. *Architecture of the implemented prototype.*

over HTTP. This tells the service consumer what the requests and responses are for the service provider.

- The service consumer uses the provided OWL-S to send a request to the service provider.
- The service provider supplies the expected response to the service consumer.
- The service consumer displays and accesses the returned WFS and WMS services using the web browser through the Internet.

Some Experimental Results

The implemented prototype is essentially a collection of ontology-based web services. Services communicate with each other in the system. It supports some fundamental interactions of the services-oriented architecture: publishing resource descriptions so that they are accessible to prospective users (*publish*); discovering resources of interest according to some set of search criteria (*discover*); and then interacting with the resource provider to access the desired resources (*bind*). The ontology-based registry and catalogue services play the essential role of matchmaker by providing publication and search functionalities. The ontology-based catalogue service enables a requester to dynamically discover and communicate with a suitable resource provider. The implemented prototype allows users to access and share geospatial data at semantic level from distributed data sources. The following examples demonstrate accessing and sharing geospatial data in a distributed environment. Figure 4 shows that data providers can map an ontology class to a feature class in a remote WFS server (http://qingdao.cs.uwm.edu:8090/geoserver/wfs) from any place over the web. In this example, ontology class *Street* is mapped to the feature *uwm:streets*, the ontology property *zipCode* is mapped to the feature property *uwm:ZIP_CODE*, and the ontology property *type* is mapped to the feature property *uwm:STREET_TYP*. Note that *uwm* is the name-space prefix that we have selected for the web feature *streets*. Also note that *ZIP_CODE* and *STREET_TYP* are the attribute names in the original shape-file data and they represent *zip code* and *type* of streets in the real world. Without mapping them to an ontology class, only data providers can understand these semantic names, and data consumers cannot understand them. Thus, they cannot query and access the original data based on these names



ontology class: Street
feature: uwm:streets
datatype property: zipCode
domain class: Street
feature property: ZIP_CODE

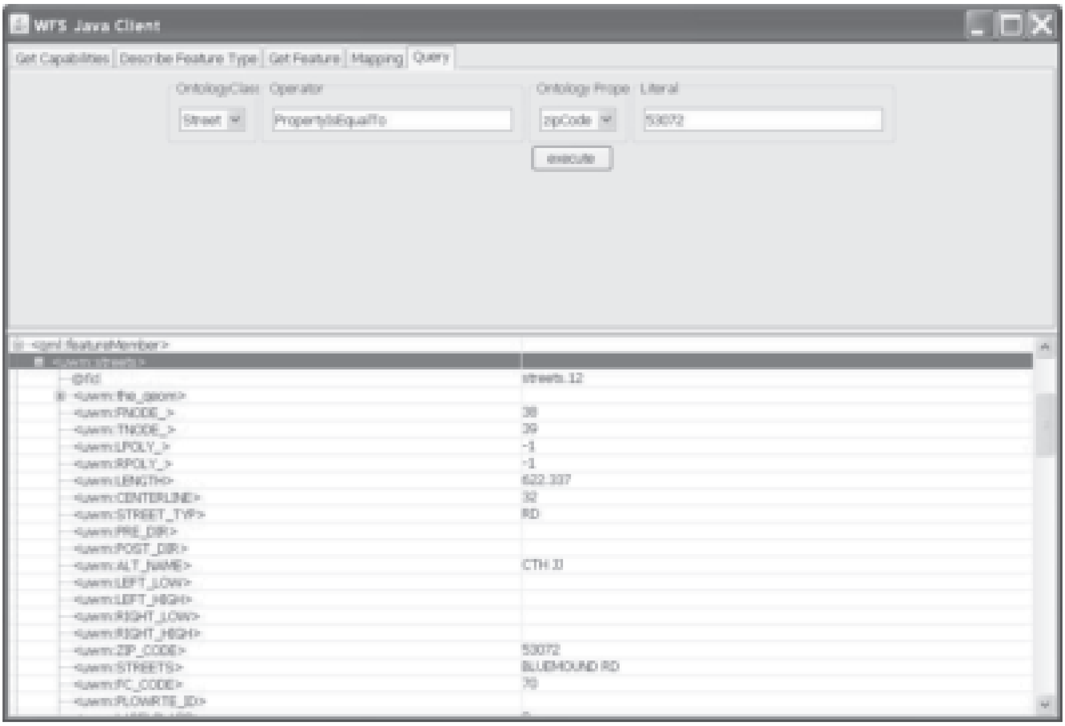
```
<wfs:FeatureCollection>
  @xmlns:gml               http://www.opengis.net/gml
  @xmlns:uwm               http://qingdao.cs.uwm.edu/geospatial
  @xmlns:wfs               http://www.opengis.net/wfs
  @xmlns:xsi               http://www.w3.org/2001/XMLSchema-instance
  @xsi:schemaLocation      http://www.opengis.net/wfs http://qingdao.cs.uwm.edu:8090/geoserver/schema...
  <gml:boundedBy>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
  <gml:featureMember>
```

Figure 4. *Map ontology class to the feature.*

in the original shape-file data. For example, they may want to access the spatial feature data of *streets with zip code 53072* over the web, but data providers only publish the original data using web feature services from a WFS server and do not register the data to the ontology-based catalogue services, therefore it will be difficult for data consumers to query and access to their interested data (*streets with zip code 53072*). However, by mapping the ontology class to the original attribute name through the ontology-based catalogue services, data consumers can obtain their interested data by directly query them over the web. Figure 5 illustrates the query results based on the ontology class *Street*, with its *zipCode* property set to *53072*. The results are the same as users would get if they use *getFeature* operation in OGC web feature services with feature *uwm:streets* and property *uwm:ZIP_CODE* equal to *53072*. Figure 6 demonstrates the graphic map of integration of different format data from disparate servers (Note: the original *bus route* data in the shape-file format are located at the data server http://qingdao.cs.uwm.edu:8090/geoserver/wfs and the original *bus stop* data in the PostGIS database are located at the data server http://zhangweb.geog.kent.edu:8080/geoserver) by invoking the ontology-based WFS and WMS services with little or no knowledge about the heterogeneous environments of the data providers (*bind*). Figure 7 shows the original data including both "bus route" and "bus stop", located in one computer and displayed using ESRI ArcGIS software. Comparing Figure 6 with Figure 7, it can be seen that the different format data published using semantic WFS and WMS services from disparate servers are matched well and there is no difference from the overlaid original data in one computer.

DISCUSSION

A framework of geospatial data sharing systems based on *Geospatial Semantic Web technologies* for heterogeneous legacy GIS is proposed in this paper in order to reduce duplication efforts among agencies and increase the benefits of using available data. Results from the implemented prototype show that the Geospatial Semantic Web based on web services offers a number of



OntologyClass: Street
Operator: PropertyIsEqualTo
Ontology Prope: zipCode
Literal: 53072

```
<gml:featureMember>
  <uwm:streets>
    @fid                      streets.12
    <uwm:the_geom>
    <uwm:FNODE_>              38
    <uwm:TNODE_>              39
    <uwm:LPOLY_>              -1
    <uwm:RPOLY_>              -1
    <uwm:LENGTH>              622.337
    <uwm:CENTERLINE>          32
    <uwm:STREET_TYP>          RD
    <uwm:PRE_DIR>
    <uwm:POST_DIR>
    <uwm:ALT_NAME>            CTH JJ
    <uwm:LEFT_LOW>
    <uwm:LEFT_HIGH>
    <uwm:RIGHT_LOW>
    <uwm:RIGHT_HIGH>
    <uwm:ZIP_CODE>            53072
    <uwm:STREETS>             BLUEMOUND RD
    <uwm:FC_CODE>             70
    <uwm:FLOWRTE_ID>
```

Figure 5. *Query based on ontology definition.*

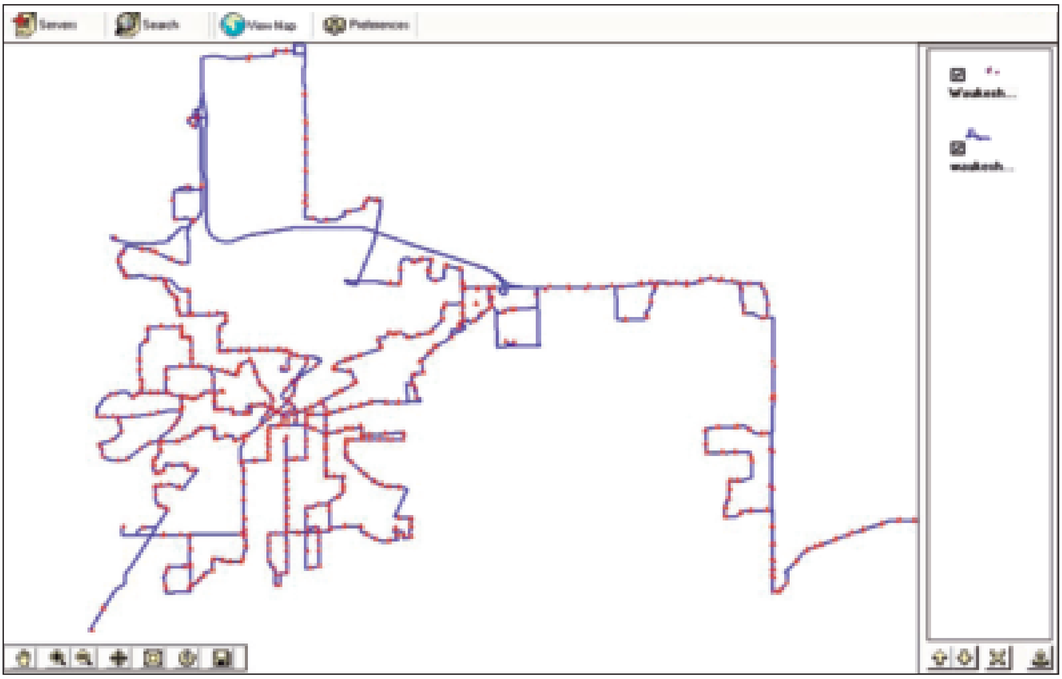


Figure 6. *Integrate different format data from disparate servers by invoking the ontology-based WFS and WMS services.*

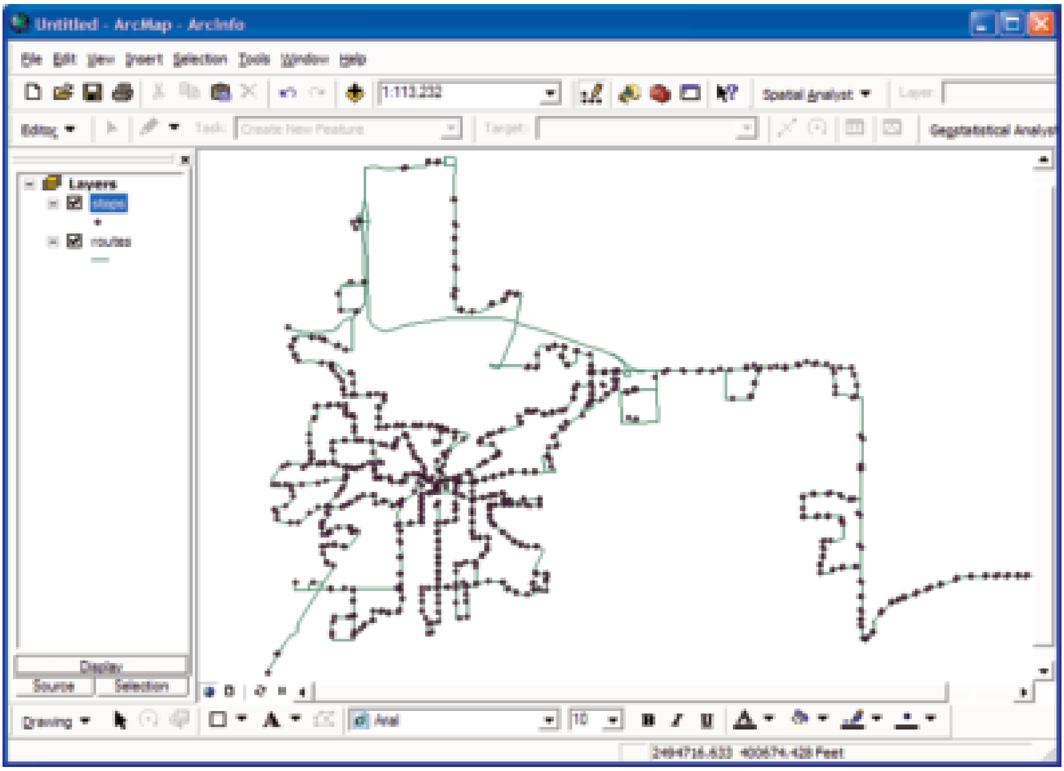


Figure 7. *Overlaid original data located in one computer using ArcGIS.*

advantages in facilitating data sharing over the web:

Firstly, it facilitates information distribution and integration at semantic level and it allows data query based on semantic content instead of keywords in metadata. With Geospatial Semantic Web technologies it will be easier to distribute geospatial data across platforms, operating systems and computer languages. Data providers can publish data dynamically on the web and data consumers can search and access the published data and web services across the web based on content instead of keywords in metadata, thereby maximizing the reuse of geospatial data. Geospatial Semantic Web makes it possible to seamlessly integrate and access heterogeneous data sources at semantic level. Users can share heterogeneous spatial data at semantic level on the fly over the web with little or no knowledge about their original data model, data storage structure, and syntactic name. Thus users can universally access spatial data at any time in any place. By making GIS resources more universally accessible, Geospatial Semantic Web makes virtual collaborations possible and has the potential to integrate GIS resources on a global scale.

Secondly, it enables flexibility through the semantic SOA. The semantic SOA, by virtue of its self-defining and standardized nature, allows GIS applications to adapt to changing environments quickly, and thus increases their flexibility and agility. With semantic SOA, it is easier for application developers to integrate heterogeneous geospatial data into their custom applications and extend legacy systems to work with new business functionalities. Geospatial web services can be *plugged into* different applications and can be combined, configured and reused on the fly to meet the ever-changing dynamics of business.

Thirdly, it reuses existing infrastructures. It allows reuse of trusted legacy GIS and applications in new ways and from different clients. Building SOAP web service interfaces and describing them with OWL-S allows legacy GIS to be accessed in a way not tied to a particular system or software vendor. Thus the vast majority of existing legacy GIS and applications which are typically closed and proprietary can interact easily with each other and the large amount of effort spent over many years in developing and enhancing legacy GIS applications will not be lost.

Fourthly, it reduces redundancy and costs. To share information, currently many organizations and businesses have to create duplicated spatial data sets in separate departments and applications, and duplication is a normal way to share information among GIS communities. Geospatial Semantic Web provides the potential to reduce the duplication by providing a distributed, dynamic, flexible and re-configurable service system over the web. The ability to reduce redundancy and reuse legacy systems in future architecture will result in significant cost savings. By reducing redundancy it also can achieve lower administration, maintenance and management costs, and potentially reduce software licensing and hardware costs. In addition, it allows a substantial saving to be realized in terms of redevelopment and integration costs by providing straightforward integration. Once an application is part of a service-oriented architecture environment, it can be accessed and integrated by any other application.

Although the proposed framework offers the aforementioned advantages, it still has several issues which need further investigation. One issue is the security issue. This issue involves confidentiality, integrity and availability. Confidentiality would provide the security element to a service, including who is authorized to use it, identity management issues, and any special security issues such as the use of encryption. Integrity implies web service messages should not be modified by any unauthorized party accidentally or deliberately during transmission. We can preserve confidentiality and integrity through authentication, authorization and encryption. Authentication and authorization ensure that web services messages should be transmitted by a properly identified sender and be read by the intended recipients. Encryption ensures that the transmitted messages are not altered or read by unauthorized users. Availability means that the web services should remain accessible to the users and thus must withstand adversaries' malicious actions, such as denial-of-service attack. For example, to add security for catalogue services in the proposed approach, we can only allow authorized individuals to publish or change the information in the registry and only the originator of the information is allowed to make changes or deletions. To secure web services, one can use the standard Secure Socket Layer and firewall based rules (Imamura *et al.*, 2004) at the transport level, while digital signatures and/or encryption may be used to protect specific parts of an XML/SOAP message at the application level. In addition, the Web Services Security Specification (Meier *et al.*, 2002; Wolter, 2004), which provides a complete encryption system, may be employed to add more security features to web services by using methods such as credential exchange, message integrity and message confidentiality.

The second challenging issue is performance. OGC web services use GML (Geography Markup Language) to encode spatial data. The network and processing overhead associated with GML is one of the major barriers to geospatial web service performance. GML is a text-based document format which makes it incredibly inefficient for network, processor and storage performances. There are three approaches to GML performance optimization: use specialized hardware accelerators, use compression approaches, and use binary GML to replace the text-based GML format. By adding more central processing units, memory and storage, the GML performance problem may be addressed at the hardware and network level. Through compression techniques such as gzip, the size of GML data files can be greatly reduced. Since the GML traffic is much larger in payload than binary-encoded traffic, there has been interest in the geospatial community toward standardizing a binary encoding scheme for GML. However, a binary GML standard might have potentially negative effects on spatial data interoperability (OGC, 2003b). In addition, the low speed of matchmaking by DL reasoning in the ontology server also affects the performance of the proposed system. The low speed of matchmaking is related to high computational complexity. Future efforts should focus on the improved algorithm to speed up DL reasoning for retrieval ontology-based geospatial information.

Except for the two major issues discussed above, other issues such as the privacy issue and the data quality issue also need further study.

CONCLUSIONS

This research proposed a framework of geospatial data sharing systems for heterogeneous legacy GIS by using geospatial semantic web technologies. A prototype has been implemented using ontology, OGC WFS, WMS and CS to share and integrate spatial data at semantic level with different data models and from disparate sources. OGC WFS and WMS are used to publish the heterogeneous spatial data. Ontologies are used to resolve the problem of differences in semantics used in different data sources. The ontology-based catalog services are employed to locate geospatial services wherever they are located, and provide information on the services they find for the user. Results from the implemented prototype showed that the proposed framework provided an environment for spatial data interoperability technically and semantically via web services and ontology. The major advantage of the proposed framework is to allow searching and access to geospatial data and services based on their content instead of keywords in the metadata. Thus the heterogeneous legacy GIS can share data and information directly across platforms, operation systems and computer languages. Therefore, it brings better reusability of existing assets and ensures existing investments to be used to their fullest extent. It provides enterprises with better flexibility in building GIS applications in an agile manner by leveraging existing GIS application infrastructures to compose new services. In spite of these advantages, several issues would benefit from further research, such as security and performance.

ACKNOWLEDGEMENT

We thank Dr Graeme Wright and three anonymous reviewers for their constructive comments on the manuscript. This research is partially supported by USA NSF grant No-0616957. Authors have the sole responsibility to all of the viewpoints presented in the paper.

REFERENCES

Antoniou, G. and Van Harmelen, F. (2003) Web Ontology Language: OWL. In Staab, S. and Studer, R. (eds.), *Handbook on Ontologies,* Springer, pp. 67-92.

Askew, D., Evans, S., Matthews, R. and Swanton, P. (2005) MAGIC: A geoportal for the English countryside, *Computer, Environment and Urban Systems*, vol. 29, pp.71-85.

Baader, F., Calvanese, D., McGuinness, D., Nardi, D., and Patel-Schneider, P. (eds.) (2003) *The Description Logic Handbook: Theory, Implementation and Applications*, Cambridge University Press.

Bishr, Y. (1998) Overcoming the semantic and other barriers to GIS interoperability, *International Journal of Geographical Information Science*, vol. 12, no.4, pp. 299-314.

Choicki, J. (1999) Constraint-based interoperability of spatiotemporal databases, *Geoinformatica,* vol. 3, no.33, pp. 211-43.

Crompvoets, J., Bregt, A., Rajabifard, A. and Williamson, I. (2004) Assessing the worldwide developments of national spatial data, *International Journal of Geographical Information Science*, vol.18, no. 7, pp. 665-689.

Cruz, I, Rajendran, A., Sunna, W. and Wiegand, N. (2002) Handling semantic heterogeneities using declarative agreements, *Proceedings of ACM GIS*, November, pp. 168-174.

Duke, A., Davies, J. and Richardson, M. (2005) Enabling a scalable service-oriented architecture with semantic web services, *BT technology journal*, vol. 23, no. 33, pp.191-201.

Fabrikant, S.I. and Buttenfield, B.P. (2001) Formalizing semantic spaces for information access, *Annals of the Association of American Geographers*, vol. 91, pp. 263-280.

Fonseca, F.T., Egenhofer, M.J., Agouris, P. and Câmara, G. (2002) Using ontologies for integrated geographic information systems, *Transactions in GIS,* vol. 6, no. 3, pp. 231-257.

GSDI (2004) Developing Spatial Data Infrastructures: The SDI Cookbook. (*http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf*), accessed July 4, 2007.

Hess, G.N., Iochpe, C. and Castano, S. (2006) An Algorithm and Implementation for GeoOntologies Integration, (*http://www.geoinfo.info/geoinfo2006/papers/p46.pdf*), accessed June 20, 2007.

Janowicz, K. (2006) Sim-DL: Towards a Semantic Similarity Measurement Theory for the Description Logic ALCNR in Geographic Information Retrieval, In Meersman, R., Tari, Z., and Herrero, P. (eds.) *SeBGIS 2006, OTM Workshops 2006, LNCS 4278*, pp. 1681 – 1692.

Imamura, T., Nakamura,Y., Makino, S. and Tamura, K. (2004) Implementation and performance of ws-security, *International Journal of Web Services Research,* vol. 1, no. 1, pp. 58-72.

Klien, E., Lutz, M. and Kuhn, W. (2006) Ontology-Based Discovery of Geographic Information Services - An Application in Disaster Management, *Computers, Environment and Urban Systems*, vol. 30, no.1, pp. 102-123.

Lutz, M. and Klien E. (2006) Ontology-Based Retrieval of Geographic Information, *International Journal of Geographical Information Science*, vol. 20, no.3, pp. 233-260.

Mansourian, A., Rajabifard, A., Valadan Zoej, M.J. and Williamson, I.P. (2006) Using SDI and Web-based System to Facilitate Disaster Management, *Computers and Geosciences*, vol. 32, pp. 303-315.

Masser, I. (2005) *GIS Worlds: spatial data infrastructures*, ESRI Press, Redlands.

Meier, J.D., Mackman, A., Dunner, M. and Vasireddy, S. (2002) Web services security *(http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnnetsec/html/SecNetch10.asp)*, accessed September 4, 2005.

OGC (2005) Web feature service implementation specification, version 1.1.0, document 04-094, (*http://www.opengeospatial.org/specs/?page=specs*), accessed September 4, 2005.

OGC (2004a) Web Map Service, Version 1.3, document 04-024 (*http://www.opengeospatial.org/specs/ ?page=specs*), accessed September 4, 2005.

OGC (2004b) OpenGIS catalogue service specification, document 04-021r2, (*http://www.opengeospatial.org/specs/?page=specs*), accessed September 4, 2005.

OGC (2003a) OWS 1.2 SOAP Experiment Report, document 03-014, *(http://portal.opengeospatial.org/files/?artifact_id=1337),* accessed July 4, 2007.

OGC (2003b) OpenGIS Geography Markup Language (GML) implementation specification, version 3.00, document 02-023r4, (http://www.opengeospatial.org/ specs/?page=specs), accessed September 4, 2005.

Peng, Z.R. and Zhang, C. (2004a) The roles of geography markup language, scalable vector graphics, and web feature service specifications in the development of internet geographic information systems, *Journal of Geographical Systems,* vol. 6, no. 2, pp. 95-116.

Peng, Z.R. and Zhang C. (2004b) GML, WFS, SVG, and the future of internet GIS, *GIS@development,* vol. 8, no. 7, pp. 29-32.

Peng Z.R. and Zhang, C. (2005) A new trend of Internet GIS development: Geospatial semantic web based on services-oriented architecture, *GIS@development,* vol. 9, no. 10, pp. 34-37.

Pundt, H. and Bishr, Y. (2002) Domain ontologies for data sharing - an example from environmental monitoring using field GIS, *Computers and Geosciences,* vol. 28, no. 1, pp.95-102.

Rajabifard, A., Binns, A. and Williamson, I. (2006) Virtual Australia: Developing an enabling platform to improve opportunities in the spatial information industry. *Journal of Spatial Science*, vol. 51, no. 1, pp. 63-78.

Smith, B. and Mark, D. (1998) Geographic categories: An ontological investigation, *International Journal of Geographic Information Science,* vol. 15, no. 7, pp. 591-612.

Tait, M.G. (2005) Implementing geoportals: applications of distributed GIS, *Computers, Environment and Urban Systems,* vol. 29, no. 11, pp. 33-47.

Williamson, I., Rajabifard, A. and Feeny, M.-E. (2003) *Developing Spatial Data Infrastructure, From Concept to Reality*, Taylor and Francis, New York.

W3C (2004) Web Services Architecture, (*http://www.w3.org/TR/ws-arch/*), accessed September 4, 2005.

Wolter (2004) Xml web services basics. (*http://msdn.microsoft.com/library/default.asp?url=/library/en-us/dnwebsrv/html/webservbasics.asp*), accessed September 4, 2005.

Zhang, C. and Li, W. (2005) The roles of Web Feature and Web Map Services in real-time geospatial data sharing for time-critical applications, *Cartography and Geographic Information Science,* vol. 32, no. 4, pp. 269-283.

Zhang, C., Li, W., Day, M. and Peng, Z.R. (2003) GML-based interoperable geographical database, *Cartography,* vol. 32, no. 2, pp.1-16.