DISTRIBUTED GEOSPATIAL INFORMATION SERVICES-ARCHITECTURES, STANDARDS, AND RESEARCH ISSUES

Liping Di

Laboratory for Advanced Information Technology and Standards (LAITS), George Mason University, 9801 Greenbelt Road, Suite 316-317, Lanham, MD 20706, USA – ldi@gmu.edu

Commission II, WG II/3

KEY WORDS: Geospatial information Services, Distributed Architectures, Standards, Knowledge Systems, Web Service, Grids

ABSTRACT:

It is estimated that more than 80% of data that human beings have collected so far are geospatial data. In order for the geospatial data to be useful, information has to be extracted from the data and converted to knowledge. However, currently it is very difficult for general users to obtain geospatial data and turn them into useful information and knowledge. In order for geospatial information to become the mainstream information, the geospatial information systems have to be able to provide the ready-to-use information that fits the individual users' needs. This paper presents the distributed geospatial services which can fully automate data discovery, access, and integration steps of the geospatial knowledge discovery process under the interoperable service framework, fully automate a rang of geo-computational services at a limited geospatial domain, and greatly facilitate the construction of complex geocomputation services and modeling. The paper discusses the service-oriented architecture, web services, Grid services, geospatial services, geo-object and geo-tree concepts and their implementation in the service environment, and the interoperable geospatial service standards. It also discusses some advanced research issues related to automatic geospatial information extraction and knowledge discovery in the web service environment, including automated geospatial data and services discovery, domain knowledge-driven intelligent geo-object decomposition, automated geospatial web service chaining, binding, and execution, and management of workflows and geospatial models.

1. INTRODUCTION

Data is a representation subject to interpretation or to which meaning may be assigned. Geospatial data is the data that can be associated with locations on Earth. Geospatial data is the dominant form of data in terms of data volume. It is estimated that more than 80% of data that human beings have collected so far is the geospatial data. Geospatial data is widely used in many aspects of socio-economic activities, ranging from environmental management to military operations. Because of importance of geospatial data, huge amounts of resource and money have been invested in collecting, managing, archiving, and distributing geospatial data. The total amount of geospatial data is approaching to exabytes and continues to grow rapidly. For example, NASA EOSDIS have already archived more than two petabytes of data, with more then three terabytes of new data arriving each day (McDonald et al., 2003).

Information is the meaning that is currently assigned to data by means of the conventions applied to these data. Knowledge is an organized, integrated collection of facts and generalizations. In order for the geospatial data to be useful, information has to be extracted from the data and converted to knowledge.

Because of the complexity of their formations and the multi-disciplinary nature of geospatial data, it is difficult to extract useful information from them without sophisticated processes. The traditional way for geospatial information extraction involves in, with help of computers, the trained specialists who are the experts on the specialized area. Because of the huge volume of the geospatial data and scarcity of the experts, most of the geospatial data have never been directly analyzed by people after the collection. On the other hand, the society is lacking of up-to-date geospatial information and knowledge for management and decision-makings. In other words, we are rich in geospatial data but poor in up-to-date geospatial information and knowledge. In order to diminish this problem, we need to develop the semantic processing that automatically derives the information and knowledge from the geospatial data in the distributed archives.

2. THE DISTRIBUTED NATURE OF GEOSPATIAL DATA REPOSOTORIES

Because of the importance of geospatial data, many public and private organizations have been engaged in collection, management, and utilizations of the data. Typically, individual organizations establish their own data centers to manage and distribute the data and products to their users. Because of the data ownership issues and the large data volumes, it is impossible to put all geospatial data into one big data center, even if within a single country. In addition, the computational resources associated with those data centers are naturally distributed.

3. STEPS AND ISSUES IN THE PROCESSES OF GEOSPATIAL KNOEWLEDGE DISCOVERY

The computer data processing in the geospatial knowledge discovery includes three consecutive steps*: A) Geoquery, B) Geodata and information assembly, and C) Geocomputation.* Geoquery is to locate and obtain data from data repositories. The geocomputation is to analyze and simulate the complex Earth system using data and information from the Geoquery. Geodata and information assembly assembles the data and information from data centers based on the needs of geocomputation.

Currently, a typical geospatial knowledge discovery process has the following steps:

1. Find a real-world problem to solve;
2. Develop/modify a hypothesis/model based on the problem;
3. Implement the model or develop an analysis procedure at local computer systems and determine the data requirements;
4. Search, find, and order the data from data providers (Geoquery);
5. Preprocess the data into a ready-to-analyze form. The preprocessing typically includes reprojection, reformatting, subsetting, subsampling, geometric/radiometric correction, etc (Geo-assembly) so that multi-source data can be co-registered;
6. Execute the model/analysis procedure to obtain the results (Geocomputation);
7. Analyze and validate the results;
8. Repeat steps 2-7 until the problem is solved.

Because of the multidisciplinary nature, geospatial data from data centers are very diverse. In many cases, the temporal and spatial coverages and resolution, origination, format, and map projections are incompatible. Data users spend considerable time on assembling the data and information into a ready-to-analyze form for the geocomputation step, even when the analysis is very simple. If datasets the user requests are not ready available at data centers, the geospatial information system cannot make the datasets for the user on-demand even if the process to make such datasets is very simple. Users have to spend considerable amount of time to order and process the raw data to produce the data products they need in the analysis. It is estimated that more than 50% of users' time is spent on the geoquery and geo-assembly steps of the geospatial knowledge discovery (Di and McDonald, 1999).

The above mode of operations in geospatial knowledge discovery assumes that the data will be acquired and input into the local computer systems for analysis. The user has to have local analysis hardware, software, and expertise in order to use the multi-source geospatial data for knowledge discovery and applications. The mode also requires significant human involvement in handling the data transactions because the analysis systems in the users' sites are normally the standalone systems and are incapable of interoperating with data systems at data centers. We call this type of mode of operations the "everything-locally-owned-and-operated (ELOO)" mode. In the past several decades, the geospatial research and applications have been all based on the ELOO mode. But this mode has significant problems:

1. Difficulty to access the huge volume of multi-source geospatial data. The process for a general user from ordering to actually obtaining the data usually takes weeks. Therefore, many applications requiring real or near-real time data can only be conducted by very few users who have access to real data sources.
2. Difficulty to integrate the multiple-source data from multiple data providers. Because users cannot get the data in user-specified form, they need to spend a lot of time and resources to pre-process the data into a ready-to-analyze form.
3. Lack of enough knowledge to deal with geospatial data. Because of the diversity of geospatial data, expert knowledge in the data manipulation and information technology is needed to handle such data. Not all users have such knowledge. In fact, many geospatial research and application projects have to hire geospatial experts to manipulate the data and operate the analysis systems. However, many potential geospatial users don't have such luxury.
4. Lack of enough resources to analyze the data. Many of current geospatial research and application projects require handling multi-terabytes of data. In order to conduct such projects, users have to buy expensive high-performance hardware and specialized software. In many cases, those resources are only purchased for a specific project and when the project is finished, the resources will be set idle.

Because of the above problems, applying geospatial data to solve the scientific and social problems is a very expensive business and only few users can afford such luxury. This is the major reason that although geospatial information and knowledge have vital scientific and social value, they are not used as wide as possible in our society.

4. MAKING THE GEOSPATIAL INFORMATION THE MAINSTREAM INFORMATION

In reality, what most users want is the geospatial information and knowledge that are ready to be used in their specific applications, rather than the raw data. However, current geospatial information systems are incapable of providing ready-to-use user-specific geospatial information and knowledge to broad user communities.

In order for geospatial information to become the mainstream information that everyone can use at will, geospatial information systems have to be able to provide the ready-to-use information that fits the individual users' needs. That means an ideal geospatial information system must be able to deal automatically with the distributed nature of geospatial data repositories and fully automate steps 2-6 of the geospatial knowledge discovery. The system has to be intelligent enough so that it can understand the description of the geospatial problem provided by the general users, ideally in nature languages, form the problem solving procedure/model automatically, figure out where the data is located and how to access them on line, run the procedure/model against the data without human interferences, and present the result back to users in human understandable forms. If such a system can be built, users only need to describe the geospatial problem accurately and examine the results. A problem that requires several months of experts' time to solve at present maybe only needs minutes or seconds to solve within such a system. Even if we cannot make such a system reality in next few years, the recent development in the service oriented architecture (SOA) and geospatial interoperability standards, as well as the advance in computer hardware and network makes the construction of geospatial information systems, which are much more capable than today's ones, possible in next few years. Such systems can fully automate steps 3-6 of geospatial knowledge discovery. Even with such a system, scientists and engineers can focus more on the creative process of hypothesis generation and knowledge synthesis rather than spending huge amount of time on those tedious data preparing tasks. The system will also greatly facilitate the construction of complex geocomputation services and modeling.

5. THE SERVICE ORIENTED ARCHITECURE AND DISTRIBUTED SERVICES

One of hot research topics in the E-business world is to enable the interoperable business services at the network environment. Currently, there are many individual standalone business services available over the Internet. However, it is impossible for individual standalone services to meet all service requirements of many users. However, such information requests could be met by dynamically chaining multiple services provided by single and multiple service providers. The service-oriented architecture (SOA) recognizes this and tries to construct a distributed, dynamic, flexible, and re-configurable service system over Internet that can meet many different users' information requirements.

The key component in the service-oriented architecture is services. A service is a well-defined set of actions. It is self-contained, stateless, and does not depend on the state of other services. Stateless means that each time a consumer interacts with a service, an action is performed. After the results of the service invocation have been returned, the action is finished. There is no assumption that subsequent invocations are associated with prior ones. In the service-oriented architecture, the description of a service is essentially a description of the messages that are exchanged between the consumer and the service. Standard-based individual services can be chained together to solve complex tasks. The basic operations in SOA include publish, find, bind, and chain (Figure 1).

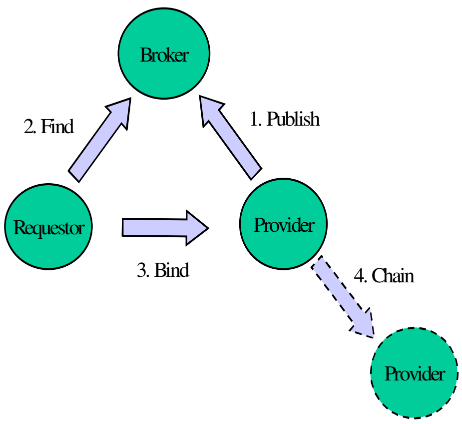

Figure 1. The basic service operations

There are three types of key actors in SOA, including the providers who provide specific services over the Internet, the requesters (users) who request information services, and the brokers who help the requestors to find the right services. When a service provider sets up a service over the Internet and wants users to use his service, he needs to *publish* the service descriptions to a broker (e.g., a registry, catalog or clearinghouse). When a requestor requests a service, the requestors and service brokers need to collaborate to *find* the right services. Service requestors describe the kinds of services they're looking for to the broker and the broker delivers the results that match the request back to the requestor. After the right service is found, a service requestor and a service provider negotiate as appropriate so the requestor can access and invoke services of the provider (*bind*). In many cases, a sequence of services must be bound together to produce the user-desired results (*chain*).

The SOA can be implemented at many different network environments. The major two include the Web and the Grid. The implementation of SOA in the web environment is called Web services. Web services are self-contained, self-describing, modular applications that can be published, located, and dynamically invoked across the Web. Web services perform functions, which can be anything from simple requests to complicated business processes. Once a Web service is deployed, other applications (and other Web services) can discover and invoke the deployed service. The real power of web services relies on

- Everyone on the Internet can set up a web service to provide service to anyone who wants—many services will be available.
- The standard-based services can be chained together dynamically to solve complicated tasks – Just in-time integration.

The Grid is a rapid developing technology originally motivated and supported from sciences and engineering requiring high-end computing, for sharing geographically distributed high-end computing resources (Foster et al., 2002, 2001; Foster and Kesselman, 1999). The vision of the Grid is to enable resource sharing and coordinated problem solving in dynamic, multi-institutional virtual organizations (Foster et al., 2001). It provides on-demand, ubiquitous access to computing, data, and services and constructs new capabilities dynamically and transparently from distributed services. The key for the Grid success is the open source middleware called Globus Toolkit (Foster and Kesselman, 1998; Globus 2004). It has become a de facto standard for all major Grid implementations. The implementation of services in Grid environment is called the Grid services.

The latest major version of Globus is Globus 3.0, which implemented the Open Grid Service Architecture (Foster et al., 2002). The fundamental concepts of services in the Grid are the same as the Web services. However, they do have differences. A Web service can be invoked by any requestors over the Web while a Grid service can only be invoked by the requestors within the virtual organization. The web service practices also have been extended in Grid to accommodate the additional requirements of Grid services, including

- Stateful interactions between consumers and services
- Exposure of a web service's "publicly visible state"
- Access to (possibly large amounts of) identifiable data
- Service lifetime management

Currently the Grid and Web communities are converging through the Web Service Resource Framework (WSRF) (GGF, 2004).

In order for SOA to work, interoperability standards related to all aspects of service operations are needed. The major international bodies setting the web service standards are World-Wide Web Consortium (W3C, 2004) and Organization for the Advancement of Structured Information Standards (OASIS, 2004) while the body setting the Grid service standards is the Global Grid Forum (GGF, 2004). The major standards related to services are shown in Figure 2.

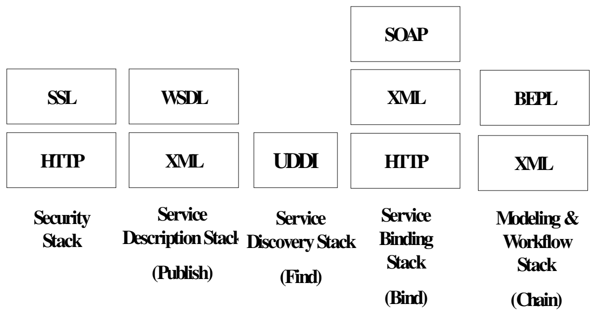

Figure 2. The major web service standards

6. DISTRIBUTED GEOSPATIAL SERVICES

Geospatial services are the services that handle the geospatial data and information (Di, 2004a). There are three major types of geospatial services: The data services, the value-add processing/client services, and the brokerage services. The data services provide the data to requestors. In the service concept, a requestor may be an application client that interacts with the human users directly or may be a middleware service that provides processing services to the data.

The distributed geospatial services under SOA have been discussed by Di and McDonald 1999, Di 2004a, and OGC (Lieberman, 2003). First, we consider a granule of geoinformation (either a dataset, a query result, or geocomputation output which describes some aspects of Earth) to be a *geo-object*, which consists of data itself, a set of attributes (metadata), and associations with a set of methods (transformation and creation methods) that can operate on it. A geo-object stored at a data center is an *archived geo-object*. All geoinformation and knowledge products are derived from archived geo-objects. Thus, from object point of view, all processes for geo-information/knowledge discovery are the processes of creating new geo-objects from existing ones.

If we consider a user request is a user-defined geo-object, or called *user geo-object*, the object is either an archived geo-object in a data archive or can be derived by executing geo-processing algorithm (e.g., unsupervised classification) with a set of input geo-objects. An input geo-object, if not exists in an archive, can be further derived by executing a geo-processing algorithm with a set of input geo-objects and so on. The decomposition process will construct a geospatial process workflow tree, which we call a *geo-tree*. The construction of a *geo-tree* is a geospatial modeling process; and the geo-tree itself is a geospatial model that contains the knowledge of a specific application domain. With the geo-tree, we know how to produce the user-object although the object does not really exist in any archives. We call such geo-object the *virtual geo-object*. In fact, any sub-tree in the geo-tree is a virtual geo-object. Since a geo-tree only captures the workflow, not a specific product, it represents a type of geo-objects that it can produce, not an instance (an individual dataset). The virtual geo-object can be materialized on-demand for users when we have all required methods and inputs available. When user requests such an object, the user has to specify the geographic location, time, format, etc. Those specifications will instantiate the virtual geo-object. By propagating the specifications down to each node of the geo-tree, the whole geo-tree is instantiated. This process is called *instantiation of geo-tree*. Only by doing the instantiation we can know if the virtual geo-object can be materialized because in many cases the required archival geo-objects may be not available for the users-specified geographic region and conditions. After the instantiation, the geo-tree is executable in the system and the virtual geo-object can be produced. The production process is called *the materialization of virtual geo-object*, which will produce an instance of the virtual geo-object.

The geo-object and geo-tree concepts can be implemented under the web service framework for automatically deriving user-requested geospatial information and knowledge. The geo-processing algorithm (method) in each node of a geo-tree can be implemented as a web service, which is called a *geospatial web service module*. The algorithm may only take care of a tiny step of overall geo-processing or may be a large aggregated processing. However, the service should be well defined, and has a clear input and output requirements, and can be executed independently. All those service modules can be reused in constructing different geospatial models. If we have enough elementary service modules available, we can construct any complex geospatial models by chaining those service modules together.

A *service chain* is defined as a sequence of services where, for each adjacent pair of services, occurrence of the first action is necessary for the occurrence of the second action. When services are chained, they are combined in a dependent series to achieve larger tasks. From service point of view, a geo-tree is a complex service chain. The construction of geo-tree is a service-chaining processing. There are three types of chaining defined in ISO 19119 (ISO, 2001):

- User-defined (transparent) – the Human user defines and manages the chain.
- Workflow-managed (translucent)– the Human user invokes a service that manages and controls the chain, where the user is aware of the individual services in the chain.
- Aggregate (opaque) – the Human user invokes a service that carries out the chain, where the user has no awareness of the individual services in the chain.

The geo-object and geo-tree concepts are currently being implemented in the GeoBrain system under the web service framework (Di, 2004a). The system uses the Open GIS Consortium (OGC) standards for the data finding and access, and OGC and W3C standards for the web services. It also leverages the new and ongoing development in the knowledge representation and management, especially the workflow management.

GeoBrain is an open, interoperable, distributed, standard-compliant, multi-tier web-based geospatial information services and modeling system. This system can shorten the time required for geospatial information extraction from weeks to just *minutes or seconds*. At the back-end, the system accesses the multi-petabytes of remote sensing and other geospatial data available at NASA and other geospatial data centers. At the front-end, end-users are able to find and obtain geospatial products in the form that exactly matches their requirements so that the products are ready for integration and analysis. In addition to obtain the geospatial data and information with great easiness, the system also provide an interoperability framework and architecture, allowing end-users to develop individual geospatial web-service modules and to use these modules and those provided by GeoBrain for constructing complex web-executable geospatial model. By executing the model on the system on-line against any subsets of the petabytes of geospatial data, end-users obtain not the raw data but a solution to their scientific questions expressed in the model. Through the proper peer review, user-developed modules and models can be plugged into the system operational services. To other end-users, the newly available web-executable models represent types of geospatial products available although the products will only be produced when users request them. These models and modules can be reused to construct even more complicated geospatial models. This accumulation of knowledge through sharing and reuse of geospatial process models make the system evolvable and increasingly capable with time. The architecture also makes the idea of interoperable systems developed by a community for the community implementable. The detail description of GeoBrain architecture and components can be found at Di, 2004. The same concepts are also being implemented in the Grid environment in another project (Di, et al., 2003; Di and McDonald, 2004; Di 2004b,c).

7. THE GEOSPATIAL SERVICE STANDARDS

It is envisioned that in the near future, there are many independent geospatial data and service providers distributed over the web. Services needed for creating a virtual geo-object may be scattered in multiple service providers. In order for the service modules to work together through the service chaining, standards on service declaration, discovery, binding, and execution have to be developed. The *geo-tree* concept requires one service's output to be the input of another service. Therefore, standards on service chaining are required. Those standards and interfaces provide the *common service environment* for service providers to deploy standard-based interoperable web services. In addition, a service may require inputs from multiple data providers. A *common data environment,* which provides standard interfaces to the data provider's archives, is required.

The common data environment is a set of standard interfaces for finding and access data in diverse data archives, ranging from small data providers to multiple-petabyte NASA EOS data archives. The environment allows geospatial services and value-added applications to access diverse data provided by different data providers in a standard way without worrying about their internal differences in handling the data.

The interface standards for the common data environment are OGC Web Data Services Specifications, including Web Coverage Services (WCS) (Evans, 200, Web Feature Services (WFS) (Vretanos, 2002), Web Map Services (WMS) (de La Beaujardière, 2002), and Web Registries Services (WRS) (Reich, 2001). The specifications allow seamless access to geospatial data in a distributed environment, regardless of the format, projection, resolution, and the archive location. The OGC WCS defines web interfaces for accessing on-line multi-dimensional, multi-temporal geospatial data in an interoperable way. Coverage data include gridded geospatial data and remote sensing images. OGC WFS defines web interfaces for accessing feature-based geospatial data. WCS and WFS together cover all geospatial data. They form the foundation for OGC web-based interoperable data access. OGC WMS defines interfaces for assembling maps from multi-sources over Web. A WMS server normally converts data to a visualized form (map) based on requirements from the requestor. OGC WRS defines web interfaces for finding data or services from the registries. The OGC technology allows requestors to specify the requirements for the data they want. An OGC compliant server has to preprocess the data on-demand based on the requirements and then returns the data back to requestors in the form specified by them.

W3C and OASIS are the leading international organization for developing general web service standards (W3C, 2004; OASIS, 2004). In the geospatial area, Open GIS Consortium (OGC) is the only international organization dedicated to develop geospatial web service implementation standards based on ISO, FGDC, INCITS, and other standard-setting organizations' abstract or content standards as well as technology standards developed by W3C and OASIS. OGC standards are fully tested at various geospatial environments. OGC specifications are widely used by geospatial communities for sharing data and resources and are becoming ISO standards. In the geospatial web service area, OGC is modifying and extending W3C standards for the geospatial web services through the OGC web service initiatives (OWS2, 2004).

8. RESEARCH ISSUES

The current implementation of service-based distributed geospatial system is mainly concentrated on enabling the user-defined and workflow-managed service chains for geospatial modeling and knowledge discovery. The technology for enabling the opaque chaining, which requires artificial intelligence for automatically construct the service chain based on user's query, are not mature yet. More research in this area is needed. The key technical questions to be answered through research include:

1. How to automatically decompose user's query (user object) to construct the geo-tree based on distributed data and service catalogs?
2. How to represent the geo-trees in computer understandable and executable workflows?
3. How to manage, share, and reuse geo-trees that represent the geospatial knowledge of a specific domain?
4. How to execute the geo-tree at the distributed web service environment automatically to derive the product that exactly meets the user's query.

Answering the first question requires domain knowledge to understand relationships and constraints among the data objects and the services that provide analysis, manipulation, transformation, etc., of that data. In the geo-object concept, if a user object is not at the archive or does not exactly match users' requirements, the geospatial information system will automatically create the object for users from the archived geo-objects through a set of dynamically chained services.

In the geo-knowledge discovery process, the geo-data and information assembly services involve in data reduction and transformation services. Such services will not change the meaning of the data they represent. Services falling into this category include data subsetting, subsampling, reformatting, geometric correction, radiometric corrections, etc. Those services are common to most geospatial analysis, data mining, and feature extraction processes. The rules for chaining such services together to derive users products are quite simple and universally accepted. Therefore, fully automating the geoquery and geo-assembly steps of geo-knowledge process is quite possible in the near term.

The geocomputation process step will derive information and/or knowledge from a set of archived objects through a set of domain-specific geospatial services. If a user requests a geo-object not ready available in archives, significant domain knowledge is needed to derive the object. For example, a military analyst is doing the battlefield traversability study. The analyst may ask the intelligent geospatial system which areas in the battlefield are traversable to a specific type of military vehicles if there are rainfalls in the next three days. To answer such kind of "what if" questions automatically by a geospatial information system requires the system to have the capability to access to all available source data and all required service modules. It also requires the system to have significant amount of domain knowledge and modeling capabilities (intelligence) for automatically constructing and executing a geospatial process models (geo-tree).

All other three technical questions are also concerned by OGC web services, the semantic Grid services, and Semantic Web communities. Therefore, technologies, especially those developed by OGC web service initiatives, should be tested in the realistic data environment. In addition, there has been work in the Semantic Web community to apply ontology ideas developed in the AI community to various aspects of Web Services and Web information search and manipulation. The goal of the emerging Semantic web services is to provide the mechanisms to organize the information and services so that human queries may be correctly structured for the available application services (the model components and data) in order to "automatically" build workflows for specific problems. That is, automatically determine the correct relationships between available and characterized data and services to generate the process models (geo-trees) and convert them to executable workflows to provide the "answers' to "what if" questions. From this point of view, research on distributed geospatial service shares the same goal as the semantic web services. The only difference is our research will deal with geospatial problems in particular.

The approach in Semantic web is to work with the Artificial Intelligence community to provide a set of layered extensions to XML in order to build up to an ontology language and tools. A set of candidate technologies includes Resource Description Framework (RDF), Resource Description Framework Schema (RDFS), and Web Ontology Language (OWL) (W3C 2004). The development in the semantic web could substantially enhance the automatic knowledge discovery from multi-source diverse geospatial data and information.

9. CONCLUSIONS

The geospatial information and knowledge are valuable to the daily socio-economic activities. The ready availability of and easy access to geospatial information and knowledge are the keys for making the geospatial information the mainstream information. The approach and system presented in this paper for automated information extraction and knowledge discovery under the interoperable distributed web service framework are very promising although significant challenges on the full automation of geospatial information extraction and knowledge discovery still exist. Although prototype systems built on this service framework, such as NWGISS and GeoBrain, has shown the significant improvement on users's discovery, access, and uses of geospatial data, information, and knowledge for their applications, many other issues, such as scalability, security, and reliability, need to be further investigated.

REFERENCE

de La Beaujardière, J. editor, 2002. Web Map Service Implementation Specification, Version 1.1.1, OGC 01-068r3. http://www.opengis.org/techno/specs/01-068r3.pdf.

Di, 2004a. GeoBrain-A Web Services based Geospatial Knowledge Building System. In Proceedings of NASA Earth Science Technology Conference (ESTO 2004). Palo Alto, CA. June 22-24, 2004. CD-ROM, 8p.

Di, L., 2004b. "The Geospatial Grid." In S. Rana and J. Sharma eds, *Frontiers of Geographic Information Technology*, Springer-Verlag. (In press).

Di, 2004c. The Development of Geospatially-enabled Grid Technology for Earth Science Applications. In Proceedings of NASA Earth Science Technology Conference (ESTO 2004). Palo Alto, CA. June 22-24, 2004. CD-ROM, 8p.

Di, L., and K. McDonald, 2004. "The NASA HDF-EOS Web GIS Software Suite (NWGISS)." In Qu, J. etc eds, *Earth Science Satellite Remote Sensing*. Springer-Verlag. (In press).

Di, L., A. Chen, W. Yang, and P. Zhao, 2003. "The Integration of Grid Technology with OGC Web Services (OWS) in NWGISS for NASA EOS Data," *Proceedings of the Eighth Global Grid Forum* (GGF8), 24-27 June 2003, Seattle, WA, USA. 10p.

Di, L., and K. McDonald, 1999 "Next Generation Data and Information Systems for Earth Sciences Research", in *Proceedings of the First International Symposium on Digital Earth*, Volumn I. Science Press, Beijing, China. p 92-101.

Evans, J. eds, 2003. Web Coverage Service Implementation Specification, OpenGIS® Project Document OGC 03-065r6, http://www.opengis.org/docs/03-065r6.pdf

Foster I., C. Kesselman, J.M. Nick and S. Tuecke, 2002. The Physiology of the Grid: An open Grid services architecture for distributed systems integration. Open Grid Service Infrastructure WG, Global Grid Forum. http://www.globus.org/research/papers/ogsa.pdf.

Foster I., C. Kesselman and S. Tuecke, 2001. The Anatomy of the Grid – Enabling Scalable Virtual Organizations. Intl. J. of High Performance Computing Applications, 15(3), 200-222.

Foster, I. and C. Kesselman, editors, 1999. The Grid: Blueprint for a Future Computing Infrastructure. Morgan Kaufmann Publishers.

Foster I. and C. Kesselman 1998. The Globus Project: A Status Report. *Proc. IPPS/SPDP '98 Heterogeneous Computing Workshop*, pp. 4-18. ftp://ftp.globus.org/pub/globus/papers/globus-hcw98.pdf

GGF, 2004. The WS-Resource Framework. http://www.globus.org/wsrf/.

Globus, 2004. Globus homepage, http://www.globus.org

ISO, 2001. ISO 19119:Geographic Information – Services. http://www.isotc211.org.

Lieberman, J., eds, 2003. OpenGIS Web Service Architecture. http://www.opengis.org/docs/03-025.pdf

McDonald, K., L. Di, and S. W. Wharton, 2002. "Strategic Evolution of NASA ESE Data Systems (SEEDS)." In Chen and Jiang, eds, *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. XXXIV, Part 2, Commission II, pp 327-329.

OASIS, 2004. Organization for the Advancement of Structured Information Standards (OASIS) homepage. http://www.oasis-open.org/home/index.php.

OWS2, 2004. OGC Web Service Initiatives, http://ip.opengis.org/ows2/

Reich, L. editor, 2001. Web Registry Server Discussion Paper. OpenGIS Project Document 01-024r1. http://www.opengis.org/techno/discussions/01-024r1.pdf

Vretanos, P. A. editor, 2002. Web Feature Service Implementation Specification, Version 1.0.0 OGC 02-058. http://www.opengis.org/techno/specs/02-058.pdf.

W3C, 2004. World-Wide Web Consortium (W3C) homepage, http://www.w3c.org