The GIS Web Portal: Beyond Data Services

Tong Zhang [1,2], Keith C. Clarke [2], Qingfeng Guan [2]

[1] Department of Geography, San Diego State University

[2] Department of Geography, University of California, Santa Barbara

Tong Zhang
Department of Geography, San Diego State University
5500 Campanile Drive
San Diego, CA 92182-4493
Tel: (805) 637-6129
Email: zhangt@rohan.sdsu.edu

Keith C. Clarke
Department of Geography, University of California, Santa Barbara
Santa Barbara, CA 93106-4060
Tel: (805) 893-7961
FAX: (805) 893-3146
Email: kclarke@geog.ucsb.edu

Qingfeng Guan
Department of Geography, University of California, Santa Barbara
Santa Barbara, CA 93106-4060
Tel: (805) 893-2705
Email: guan@geog.ucsb.edu

The GIS Web Portal: Beyond Data Services

Tong Zhang [1,2], Keith C. Clarke [2], Qingfeng Guan [2]

[1] Department of Geography, San Diego State University

[2] Department of Geography, University of California, Santa Barbara

Abstract

The GIS web portal usually functions as a web-based gateway to integrated and distributed geographic resources. Although current GIS web portals are gaining popularity among GIS users and are of great significance for building distributed Internet GIS, they remain weak in terms of analytical and modeling components. We discuss the possibilities of integrating state-of-the-art technologies which include Grid computing, Web Services, OGC (Open GIS Consortium) interoperability standards and the Semantic Web. Emphasis is given to the exploration of technical details of how we can combine these components together to build comprehensive GIS web portals that are capable of supporting advanced GIS analytical services. Prior and present research from the Geography and Computer Science communities is reviewed. A new methodology framework is introduced and based on it, a simple demo is presented which not only tests some of our initial ideas but also illustrates the challenges Internet GIServices developers have to face when implementing powerful yet intelligent Internet GIServices in the future.

Keywords: GIS web portal, Grid computing, OGC, Internet GIServices

1. Introduction

The development of the GIS web portal has received increasing attention in applications of Internet GIServices. A Web portal has proven itself to be an efficient way to disseminate geospatial data and information. Recent GIS web portal projects largely focus on commercial and governmental applications (Tait 2005). Examples are National Geographic MapMachine, ESRI's Geography Network and the European Geo-Portal. Basically, these GIS web portals usually function as a web-based gateway to distributed geographic resources, particularly GIS datasets. By exploring the GIS-related metadata online, users can locate, view and download the GIS data of interest. With current Internet GIS technologies, geospatial data and web-based mapping services are no longer technical challenges. Although current GIS web portals are gaining popularity among GIS users and are of great significance for building distributed Internet GIS, they remain weak in terms of analytical and modeling components. The fact that these functions are largely absent from GIS web portals can be explained by two technical barriers: performance overhead introduced by more complicated demands from Geocomputation and

Geosimulation; and interoperability issues introduced by the different proprietary GIS data and functions. The complicated Geocomputation demands significantly undermine the communication speed of Internet GIS, making the interaction between user and system difficult to maintain. Distributed GIS data and analytical components also find themselves hard to connect and interact due to interoperability problems.

Can GIS web portals go beyond GIS data services and offer advanced Geocomputation and Geosimulation services online via a GIS web portal? Before we can say "yes" firmly, it is imperative to carefully review current computing technologies to see if it is possible to implement advanced GIS web portals with analytical components that are capable of handling complex geospatial problems. Can recent emerging computing technologies, such as Grid computing, Web Services, OGC (Open GIS Consortium) interoperability standards and the Semantic Web, be integrated to realize this possibility? These technologies are supposedly helpful in working around the two barriers (performance and interoperability concerns). The GIS web portal can definitely take full advantage of these new technologies to deliver powerful yet intelligent Geocomputational services. In this article, emphasis is given to the exploration of technical details of how we can combine these techniques together to build comprehensive GIS web portals that are capable of supporting advanced GIS data and analytical services. We will review related research particularly in a geospatial context. A new methodology framework is introduced and based on it, a simple demo is presented to prove that our initial idea is feasible

The rest of the paper is organized as follows: section 2 briefly introduces present GIS web portal techniques and applications. Section 3 reviews several critical technologies pertaining to building advanced GIS web portals. Section 4 elaborates on our approach to implement a web portal based problem solving environment given complex geospatial problems. Section 5 reports on a simple experiment as a pilot study to test part of our methodology. The last section closes with brief discussion and future research plans.

2. GIS Web Portal

The Web portal has been chosen as a popular way to integrate geospatial resources. A Web portal offers a centralized and uniform interface to access the distributed and heterogeneous resources and services. Web portals have been widely implemented in many commercial websites to offer personalized web-based services such as web-based email services, personalized news, calendars etc. In the GIS community, we also have a fast growing use of web portals to deliver GIServices. Tang and Selwood (2005) classify the spatial portals as three categories: "application portals, catalog portals and enterprise portals". There are many examples and demonstrations for each category. The efficiency has been proved in many geoportals as Tait (2005) introduced and most of them fall into the categories of "application portals" and "enterprise portals". Apart from these commercial and educational GIS web portals, the scientific community adopts the web portal strategy as well. The NASA's Earth-Sun System Gateway presents a demo of accessing and visualizing geospatial datasets remotely (NASA ESG 2006). The scientific GIS web portals might incorporate functions in both “application portals”, “catalog portals” and “enterprise portals”. As a data-oriented discipline, GIS researchers are really concerned about geospatial data. The data interoperability issues thus need to be carefully handled. \ A noticeable feature of GIS web portals is that normally they don’t provide as many analytical tools as other scientific portals do. They usually focus on the geospatial data services like data searching and viewing. Users have to download the data to local machine and analyze with desktop GIS software. This is partly due to the organizational issues and partly because of the technical challenges to develop interoperable GIS analytical tools online.

As for Grid-enabled GIS web portal, there are very few demonstrations dedicated to geographic problems in spite of many ongoing projects in Geosciences. GISolve is developed at University of Iowa (GROW 2006) to work as a problem solving environment for spatial statistics computation. It presents an applicable developing strategy to build an entrance to Grid-based GIS analysis tools. According to Wang (2006), GISolve is built upon a three-tier architecture of web client, portal server and Grid resources. Some extensions are made to make the Globus Toolkit and the Jetspeed port server support geographical analysis. This research can be seen as a pioneering within geography as it exhibits particularities of geographic analysis and implements a problem solving environment demo.

3. New Opportunities for Internet GIServices

We are currently at a point where Internet GIServices have achieved tremendous successes in terms of offering web mapping and geospatial data services. Yet, more opportunities are awaiting us to continue to promote Internet GIServices to higher level where they can be more capable of assisting people to make smart spatial decisions. These opportunities come from recent breakthroughs within the Internet GIService community as well as newly emerging computing technologies.

3.1 Recent Breakthroughs in Internet GIServices

To summarize the recent breakthroughs of Internet GIServices, there are basically four aspects to be addressed: open source developing strategy; significant trend toward adopting OGC standards; revolutionarily improved performance and availability; and ever-increasing attentions to Internet GIServices from the general public.

In addition to commercial web mapping programs, open source Internet GIServices have gained a tremendous rise in popularity in GIS community. We can easily get access to many open source software for free. They cover from the Internet GIServer (MapServer), GIS database management system (PostGIS), to GIS analysis tools (GeoTools). The literally no-cost develop strategy plus free source code provide a non-proprietary solution for Internet based GIS applications (Anderson and Moreno-Sanchez 2003). The recent creation of Open Source Geospatial Foundation could be seen as a major event for joint efforts from previously uncoordinated endeavors (Open Source Geospatial Foundation 2005).

When talking about open source Internet GIS, it is impossible to avoid the OGC specifications. They are actually the de-facto standards for the entire open source GIS community and many GIS companies are announcing their support for OGC standards. To address interoperability issues, the OGC specifications, especially the Web Map Server Specification, Web Feature Service Implementation Specification, Web Coverage Service Implementation Specification, Catalog Service Specification, Geography Markup Language, are the most adopted standards. Building OGC-compliant Internet GIServices is almost an indispensable part of current Internet GIServices programs.

The release of Google Map series products (Google Map, Google Map API, Google Earth) as well as similar competitors (e.g. Windows Live Local) made a considerable contributing to the almost totally new scenario of Internet GIServices. The unprecedented high performance and data availability undoubtedly appeals to the general public. Vivid satellite images plus smooth zooming and panning offer any geographic detail from global to street level. The amazing performance is realized through a bundle of techniques: Telcontar's Drill Down Serve as server platform; Rich Map Format as new data format; and AJAX (Asynchronous JavaScript and XML) as client/server communication mechanism.

All of the above changes have popularized Internet GIS and transformed thousands of people into Internet mapping fans. After experiencing all these revolutionary changes, there are millions of teachers, students, government officials, entrepreneurs and laymen who use current Internet GIServices will come to realize the charm of Internet GIServices. Promotion of awareness by the general public (Tsou 2005) will inevitably contribute to the future big leap of Internet GIServices when new ideas and applications are emerging or coming up from the collective imagination.

3.2 Supporting Technologies to Empower A GIS Web Portal

There are several powerful supporting techniques can be used to empower the GIS web portal: Grid computing, semantic web, web services et al. Combined with these technologies, GIS web portals can offer Internet GIServices which are powerful, intelligent, and reliable and can be accessed on a very convenient basis.

3.2.1 Grid Computing

Grid technologies have been rapidly evolving since mid1990s. They are primarily concerned with the issues on the integration of large-scale computational resources and services (Baker et al 2002). The increasing diversity of computational and human resources created the "Grid problem" which requires dynamical resource sharing mechanism between "Virtual Organizations" (Foster et al 2001). "Virtual Organization" is in particular a ground-breaking concept for crossing the administrative and institutional boundaries for resource and services sharing. Researchers in computer science community have been focusing on the low-level Grid infrastructure construction. One famous example is the Globus Toolkit (Foster and Kesselman

1997). This Toolkit provides a standard Grid framework and resource coordination mechanisms. The most important motivation of Globus Toolkit is its standardized and open Grid protocols (Foster et al. 2001). Many undergoing Grid projects across the world are based on the Globus Toolkit. The newest release of Globus Toolkit is version 4.0.1 which is based on OGSA (Open Grid Services Architecture) standards and incorporates the newly proposed WSRF (Web Service Resources Framework). OGSA can extend the capability of Web Services into the Grid computing framework while WSRF allows users to manage the state of Web Services (Globus Alliance 2006). The Globus Grid Forum are working with other organizations to design standards and specifications for Grid protocols on data management (GASS, Global Access to Secondary Storage and GridFTP), resource management (GRAM, Grid Resource Allocation and Management Protocol.), security (GSI, Grid Security Infrastructure.) and information services (MDS, Monitoring and Discovery Services). Many application services could be created to solve the problems in all kinds of disciplines. According to Globus Alliance, the successful Grid test bed examples include the applications in blood flow simulation (Brown University), high energy physics (CERN), earthquake simulation (Southern California Earthquake Center), and magnetic fusion experiments (National Fusion Collaboratory), Hurricane visualization(TeraGrid), gravitational effects simulation and climate data visualization (ESG) (http://www-unix.globus.org/alliance/impact/).

3.2.2 Semantic Web

Semantic Web is an initiative to facilitate web-based data sharing within the global network system. By providing a better definition of web-based data and services, large scale data sharing and reuse become possible (Berners-Lee et al. 2001). A series of relevant standards and specifications have been defined by W3C (World Wide Web consortium) aiming at promoting the application of semantic web technologies (W3C 2006). Among them, RDF (Resource Description Framework) and OWL (Web Ontology Language) are the most important components. RDF is designed to organize web information into triple terms (subject, predicate, and object) for easier data sharing and retrieval (http://www.w3.org/RDF/). The "Uniform Resource Identifier" can be used to mark globally unique resources. When described in XML, RDF style resources can be automatically processed thus making the intelligent web data sharing possible. Web Ontology Language (OWL) is also proposed to handle terms and relations in semantic web by defining terminology used for specific contexts and properties in terms of classes and relations (http://www.w3.org/TR/owl-guide/). In addition to the reorganization of online content, structuring web-based services should not be ignored. OWL-S is a specification to help discover and use services (http://www.w3.org/Submission/2004/SUBM-OWL-S-20041122/). Based on these standards, a number of semantic web tools have been implemented for data parsing, metadata processing, ontology management and RDF/OWL formatting (e.g. Jena, RDF Gateway, Unicorn system et al.).

Geographers have been researching ontology for formal geographical representation for solving

GIS interoperability issues for a while (Smith and Mark 2001; Mark et al. 2004; Agarwal 2005). The formalization of geographic concepts, relations, categories and inference is the major goal. If this becomes true, the GISystems will benefit from the improved interoperability with universal geographic terminology system. The Ontology-driven GIS approach is proposed for testing the effectiveness of this solution (Visser et al 2002; Fonseca and Câmara 2002; Fonseca et al. 2003). Obviously, defining the ontologies for the entire geography world is impossible (Mark et al. 2004). Building hierarchical ontologies appears to be feasible to break down the entire geography conceptual space (Fonseca and Câmara 2002). Despite both the conceptual and empirical studies, the geographical ontologies seem to be still far away from the real applications.

The development of the semantic web might provide an opportunity for geographers to describe geographic space in a more efficient way. The W3C standards, plus the emerging techniques such as ontology mapping, data integration, and semantic search, if extended with geospatial components, will contribute to the building of intelligent searchable GIServices in a dynamic distributed network environment. There have been some ongoing research projects in the area of semantic geospatial services. Codex (Pike and Gahegan 2003) is a web portal with an interactive graph visualization interface for geo-spatial ontologies and knowledge management (http://flatbox.geog.psu.edu/codex). Arpinar et al. (2005) introduced their efforts on geo-spatial ontology development and supporting semantic analytics methods. Another effort toward interoperability is made by Lin and Ludaescher (2004) who implemented ontology-enabled geo-spatial data integration system based on ArcIMS, a vendor-based Internet Map Server.

3.3.3 GIS Grid

While Grid computing is dominating the HPC (High Performance Computing) world, GIS people are also attracted to the new computing paradigm. Although very few current Grid-based projects are geospatially related, these initiatives do announce a new path for GIServices. Geocomputation is mainly used by researchers of geography and other disciplines. GIServices, on the other hand, are user-oriented. The word “GIServices” is not a fancy phrase coined to help those who are scared by jargon like “Geocomputation”. Grid, which is quickly merging with SOA, offers an ideal platform for implementing GIServices which usually are concerned about performance and interoperability. However, Grid only provides general-purpose protocols which have to be modified to fit the special needs of geospatial applications. The source of these particular concerns results from the characteristics of geospatial data. These special properties include spatial-temporal dependency, reference framework, spatial heterogeneity (Miller and Han 2001), high reliance on visualization, large data volume, as well as particular scale, security, and uncertainty concerns. These special issues will in turn be reflected in the design and implementation of GIServices. All these complexities introduced by geospatial applications require more consideration in terms of scalability, reliability, resource management, graphical knowledge management and security issues.

Grid-based GIServices have to take into account a variety of issues and this can start from looking into the Grid architecture. Foster et al. (2001) describe the general Grid architecture which consists of four layers: "application, collective, resource, connectivity and fabric". At the "Fabric" layer, geospatial applications might involve remote sensing sensors, GPS receivers and other geospatial data gathering devices. These particular units will require special interfaces with other resources and GIServices. The OGC SensorWeb ™ initiative will focus on facilitating sensor-collecting, data exchanging and encoding issues (OGC 200). Geospatial metadata are another thing to be considered in this layer. The management of metadata should be able to provide easy and transparent access and enquiry mechanisms. The rich content of geospatial metadata will pose a challenge for realizing this goal. When it comes to the "Connectivity" layer, it primarily deals with Internet communication and security solutions. Given sensitive geospatial data, the problem arises when different privileges are granted to viewing and manipulating geospatial data with different resolution, detail and quality. The "scale" problem is unique for geospatial data, thus we have to come up with addition solutions besides the general mechanisms. The "Resource" layer handles individual resources. In additional to general-purpose "Information protocols" and "Management protocols", the capabilities of handling geospatial resources like metadata should be added. The "Collective" layer, which interacts with multiple resources, covers a wide range of protocols which mostly support VO operations. From the perspective of a geospatial VO, the Grid-enabled GIServices will support highly scalable operations like efficient query of geospatial resources, fast and interoperable geospatial data transferring, geospatial oriented parallel processing and powerful fault-tolerance capabilities. All of these can get hints and tips from previous and ongoing studies such as ArcIMS Metadata Server, OPeNDAP/DODS services, or extend current framework such as developing geospatial extension/libraries of MPI that optimizes the geospatial data processing and manipulation.

The geospatial VO security issues are again critical and more difficult to implement at the group-level than that at "Connectivity" layer. For geo-collaboration, the role of maps should not be underestimated and underplayed in the process of decision making such as alternative evaluation and selection. The "Application" layer resides at the top of the Grid architecture. The geospatial applications are supposed to work with other applications like workflow systems at this layer. Integration is a big issue for this layer. The interface to low-level of Grid middleware and making the applications Grid-enabled are also important. For each layer, to enrich geospatial capabilities, additional APIs need to be developed.

3.3.4 Other Emerging Techniques in Internet GIServices

There are other promising techniques that can enhance Internet GIServices. These techniques mainly deal with performance, interoperability and usability issues. In terms of performance issues for Internet GIServices, many researchers focus on the optimized organization and efficient transfer of geospatial data. For example, Yang et al. (2005) discussed several innovative web techniques to improve the performance of Web-based GIS. These techniques include pyramid techniques and harsh index, cluster and multithread, caching and dynamic data request, binary format and compression which work on both server and client side to enhance data access efficiency. Commercial web mapping products (e.g. Google Earth) also demonstrate the effectiveness of some new techniques such as AJAX (Asynchronous JavaScript and XML) which facilitate and improve the communications between clients and servers. All these web techniques can be integrated into the geospatial cyberinfrastructure architecture to achieve better Internet GIServices performance. Again, the importance of interoperability should not be underestimated since Internet GIServices are primarily distributed. In addition to the efforts stepped up for distributed data access, OGC (Open Geospatial Consortium, Inc.) is working on a standard specification of online GIServices integration framework (http://www.intl-interfaces.com/servicemodel/) with which interoperable Internet GIServices are easier to be integrated, discovered and accessed. Motivated by SOA and Web Services, GIS Web Services are widely accepted as a solution for loosely coupling GIS components in distributed environments. Proprietary commercial products start to support Web Services such as ESRI ArcWeb Services (ESRI 2006). Some researchers are working on extending Web Services to better manage GIServices (Mehmet et al. 2005).

4. A Routine-Constrained Geospatial Problem Solving Approach

4.1 Obstacles to Using Internet GIServices in the Real World

As discussed in the introduction section, a GIS web portal is an effective way to deliver geospatial data and web mapping services. However, given complex geospatial problems, currently there is no mature Internet-based GIS solution. People still feel more comfortable using desktop GIS packages to deal with the problems at hand even if this solution costs more than other alternatives. To deal with geospatial data inconsistence, integration, and complicated analysis and modeling are not easy jobs in a desktop environment, not to speak of using Internet GIServices. Obviously, using Internet GIServices has the advantages of data sharing, low costs, less training, and more users. However, when it comes to complicated geospatial problems which usually involve huge volume of geospatial data, complex modeling and intensive Geocompuation/Geosimulation, the disadvantages of Internet-based solution might outweigh the advantages. Security issues are particularly huge concerns in Internet environments (e.g., sensitive data, authentication, authorization, delegation). The lack of most common GIS analysis functions makes it impossible to conduct complicated analysis online. There where are even online GIS data/analytical services available, the automatic searching and selecting appropriate GIServices (given the criterion of performance, effectiveness, suitability and availability) are hard. Intensive Geocomputation/Geosimulation gives rise to the performance issues again although there are already techniques to handle performance issues for geospatial data online. They are proprietary techniques and cannot be used by common users. And more importantly, Geocompuation and Geosimulation are basically computing processes rather than static data storage and cannot be handled without innovated spatial database techniques. Interoperability issue, although addressed by OGC and other efforts, is still a major obstacle in dynamically building Internet GIServices especially for particular domains.

A positive side of Internet GIServices is their simplicity: a web browser is enough for fulfilling almost all tasks. To demonstrate the simplicity with complicated problems solving, an integrated and united web portal should be implemented as web-based problem solving environment. To assure users of the validity of analysis results, the GIServices have to provide a set of monitoring mechanisms to keep track of data transfer and analysis progress. To realize auditability, a global naming system for GIServices has to be established to identify every available GIService. Finally, many geospatial applications demands persistent GIServices, thus the stability would set new challenges if the remote GIServices are not controlled by users. Fault tolerance should be paid sufficient attention.

4.2 Finding a Way around Obstacles

To sum up, the major obstacles of using Internet GIServices to solve complex geospatial problems are primarily due to the ill-structure nature of these problems. Different people may consider the problems from different perspectives and this may lead to the varies of problem definition and description. These inconsistencies in turn would produce different problem solving procedures which require different geospatial data, analytical tools, and workflows. Typically to solve the problems, people have to go through a series of GIS analysis, modeling, visualization and simulation operations. They are nontrivial services. If a GIS web portal can be used to address these problems as a web-based geospatial problem solving environment, we can take advantage of the benefits of an integrated and easy-to-use interface which allow users to access not only distributed geospatial data but also complicated geospatial analytical tools. Given complex geospatial problems, we have to tackle them case by case because every single problem solving method is distinct with special spatial, time, scale concerns specific to every single problem. Traditionally, given a specific problem, there are usually a set of defined problem solving routines which vary with different users. For a defined multi-step procedure, we just follow the steps to take in input data and get the results after going through logically connected operations. From the other hand, some users prefer to conduct abductive inference and inductive learning with interactive data mining strategies. There are not fixed steps to solve a problem. This approach grants more freedom to the users who may apply their domain knowledge in tackling particular problems. We can combine the advantages of the two strategies with a routine-based user-centered analysis approach so we may not be lost in the overwhelming data mining/visualization information. At the same time, the users can apply their own domain knowledge given specific operations, data input and expected output.

As analyzed in the last section, the obstacles of applying Internet GIServices for complex geospatial problem solving are primarily associated with performance, security and interoperability concerns are the major issues which result in the lack of most common GIS analysis functions. The difficulties of searching, locating, selecting and connecting best matching GIServices make it even hard to build truly distributed GIS solutions. However, from the technical perspective, the lack of comprehensive Internet GIS solutions for complex geospatial problems can finally be solved by utilizing the new technologies discussed above. At least it looks promising and feasible. We now have many emerging techniques from computer science. Grid computing provides a comprehensive framework to offer qualities of services in secure and efficient way with global computing resources. Semantic Web can transform the web based information so they can automatically be processed by machines. This can play two roles: describing the formalization of problem solving routines and helping build intelligent GIServices search and discover mechanisms. The semantic routines make it possible to exchange different solutions among different users. Web Services are extremely helpful to make the communication between distributed software much easier. It is impossible to realize smart GIServices without them. In this context, a GIS web portal as a web-based problem solving environment would simplify the interface between researchers and complex configurations. It also can function as a collaborative environment in which people can work as virtual organizations when combining with Grid computing.

Within GIS, thanks to the recent development of Internet GIServices, we can easily get access to many open source software for free. So we don’t need to pay a lot to find the tools/programs that we can use to build comprehensive and sophisticated applications. OGC standards (WMS, WFS) have been widely adopted, which significantly facilitate the interoperability of Internet GIServices. The high performance and data availability of recent web mapping services, like Google Map, provides new solutions for geospatial data distribution (AJAX). Finally, the awareness of the general public will contribute to the future evolution of GIS web portals when more and more people are participating in this development by contributing their new ideas and technical skills. It is expected that the future GIS web portal will be powerful enough to allow common users to access geospatial data, conduct spatial analysis, run simulations, and visualize the results at even global scale. Through a simple interface, users can take advantage of any available geospatial resources including data, information, analysis tools and even data collection instruments (e.g., PDA, GPS receivers, remote sensing satellite). Without the supporting technologies, the geospatial resources will easily overwhelm the users who cannot find the best resources and services.

The complex geospatial problem solving approach can be described in the following architecture (Figure 1). The first tier is the presentation tier which is the interface between users and the underlying sophisticated distributed computing environments. The GIS web portal offers a set of tools to assist users to solve the given problems. Several views should be available for users to examine the problems from different perspectives, including a workspace view, data view, map view, logic view etc. Users can switch between these views. The workspace view provides a high-level vision of the problems in which users can have basic information, including the participating people, problem solving history, current progress, work calendar, so as to better track progress. Data view presents the data used in current project. Maps should not be overlooked in geospatial problem solving. Despite predefined workflows, maps still can be used to promote spatial thinking to explore spatial data and to produce new hypotheses. A map view should allow users to visualize the given data and intermediate Geocomputation results and directly manipulate the map to interactively operate on the data. The workflow is presented as a multi-step process which uses different diagram shapes and colors to describe different elements and operations as well as priority level. The problem granularity will be reflected in the workflow view. Users can zoom in to examine the problem at a more detailed level. Users can visually and interactively manipulate Internet GIServices as analysis modules (e.g., spatial querying, buffering, and layering, spatial autocorrelation, finding spatial patterns such as clusters and outliers). . This way user can implement their personalized analysis strategies in addition to the predefined procedure.

Logic view stores the semantics for problems at hand and will be updated when users modify the contents in other views, for example, editing the data or using new GIServices. This view is directly connected to the logic tier. This is a thin client since most of intensive computation will take place in other tiers. In this tier, task monitoring information and the Grid resource scheduling information will be provided for users to monitor current job progress, track the computing resources currently used, and take actions to interrupt the running jobs if needed. The GIS web portal problem solving environment will provide a graphical environment (which is customizable) to connect the steps into a united model with workflow style (like ArcGIS model builder). We can zoom in to see different levels of details. For example, inside a sub-task, we can see what datasets are included. We can take a look at the map or zoom in to see any details. Users can do interactive data analysis here for any given data and the systems will search the GIServices to do the analysis work. This can be combined with some black box Geocomputation or Geosimulation models which don't need prior information and produce intermediate results. If a user can easily access Internet-based distributed GIServices, they can conduct interactive geospatial data analysis as they want to test the models, adjust the parameters and make them better fit the data and problems. Since this project can be saved and shared, plus the highly interoperable geospatial semantics, the proposed GIS web portal solution can be further built into a geo-collaboration tool. Finally, it is supposed to be highly automated, the tasks can be done with minimum human intervention.

The visualization of work flows is supported by the logic tier where geospatial problem semantics are located. The Semantics will be described by geographic OWL (Web Ontology Language). Powered by this tier, the user can wrap up the workflows as their particular solutions which are delivered to their collaborators. Usually a workflow consists of multiple sub tasks. These sub tasks will in turn include smaller sub-tasks. These multi level structures will be stored in the Semantics. Users with required privilege can modify the Semantics such as adding, deleting subtasks or changing the elements for subtasks. The interfaces between previous sub-tasks and subsequent sub-tasks should be carefully defined to describe input data and output results formats. Within the sub-tasks, there will be suggested GIServices which are customizable. These GIServices can be hosted by anyone, government agencies, research institutions and private companies as well as individuals. However, they have to be OGC-compatible web services and Grid-enabled. We expect to automatically locate the most desirable GIServices given data input and expected output requirements in a specific step. The address and identifier of the located GIServices then can be returned to the representation tier dynamically to build a real-time application.

The last tier, the Grid tier, will function as the underlying infrastructure to support the high-level GIServices. It usually includes the Grid Fabric, Connectivity and Collective Layers (Foster et al. 2001). However, the data accessing, job management, failover, authorization, authentication, service discovery, and diagnostic mechanisms have to be modified to reflect the requirements of Grid-enabled GIServices. This infrastructure includes all the underlying Grid computing hardware and software. They provide a supportive environment to make high-performance geospatial analysis possible. High-speed interconnection networks, processors, and storage space will be the basic fabric elements. Coupled with Grid software, these hardware can be connected to provide aggregated manageable resources. In addition, our GIS researchers may have other concerns about adding geospatial collection devices into this Grid infrastructure. These devices will include high-resolution remote sensing systems, GPS navigation systems, mobile GIS hand-held devices, etc.

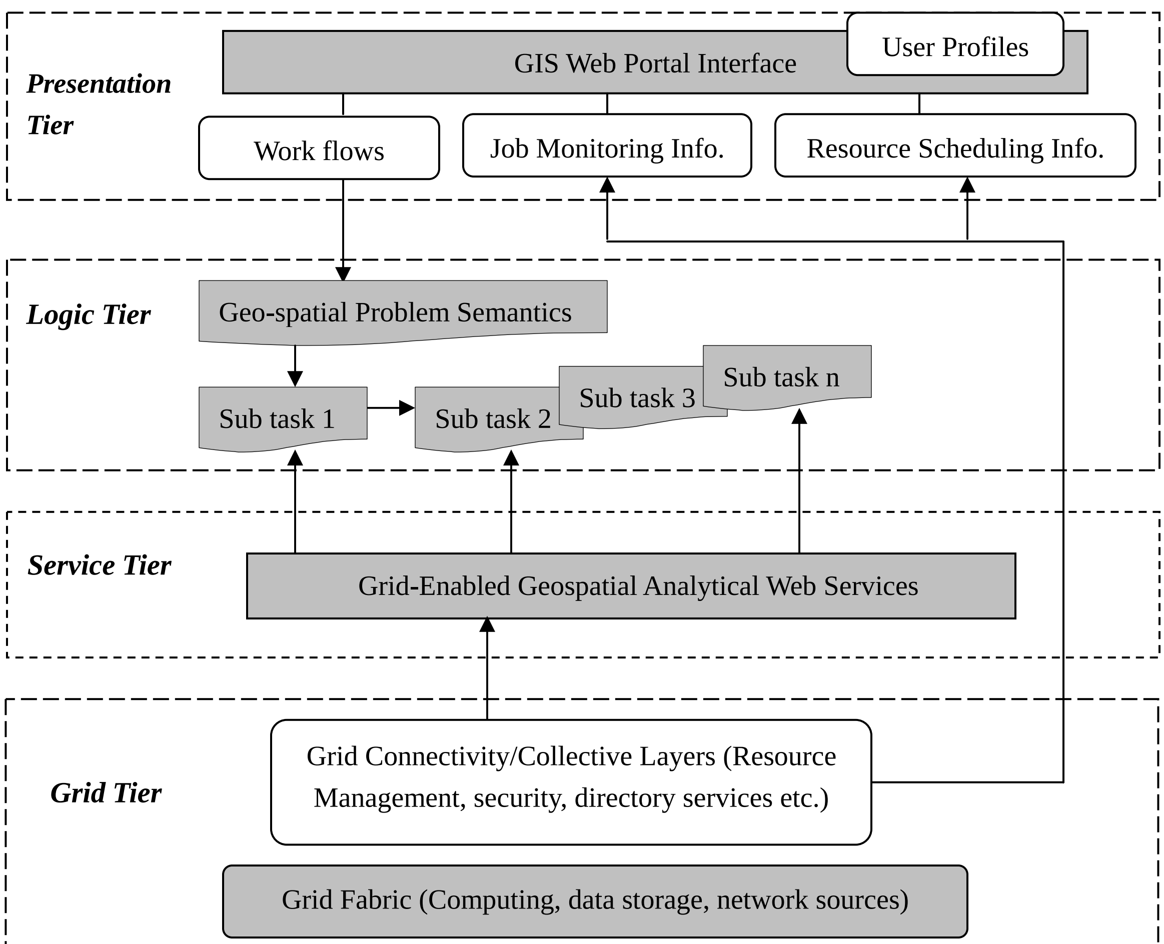

Figure 1. A Routine-Constrained GIS Web Portal-Problem Solving Approach Architecture

Given current computing and Internet GIS technologies, this architecture can already be implemented. For the presentation tier, we have a web portal framework which allows developers to customize the interface and more importantly function as containers to the run-time environments where application programs are deployed. In the logic tier, semantic languages can sued to designed and describe the workflow. The workflow engines are widely available. Particularly, Grid-enabled workflow research has produced tangible tools. The GIServices implemented in OGC standards and deployed in Grid environment have been proven to be applicable. In the Grid tier, the Grid middleware (e.g., Globus Toolkit) can be easily downloaded and installed.

The primary developing principles are as follows: (1) Our research will adopt open source developing strategy; (2) It will take advantage of grid computing with its resource coordination, security, data management, and information management mechanisms; (3) The architecture will follow SOA design principles and basically all the smart GIServices are web services; (4) The Grid-enabled GIServices will be OGC compatible; and (5) The GIService search and discover will rely on geospatial semantics.

5. Pilot Study

To test that at least part of our initial idea is feasible, we conducted a pilot study using a K-Means clustering algorithm as a GIService to be deployed in a web portal. Although the study is still ongoing, we are more confident in using current available techniques to develop intelligent analytical GIS web portals.

We implemented a K-Means clustering algorithm in Java and deployed it in Globus Toolkit as a Grid Service. Then we can develop a portlet based on the grid service. This portlet in turn will be deployed in GridSphere web portal and make available for the Internet users. The service will read the point data in GML format and produce the clustering results in GML which will be processed as SVG and visualized.

These are the tools and techniques we are using in the pilot study.

1. Globus Toolkit. It is the most popular Grid computing software and we use it to develop and deploy clustering grid service.
2. GridSphere is a Portlet-based portal framework which is highly customizable.
3. The JSR 168 portlet specifications define portlet API so the developers can develop compatible portlets for different portal containers.
4. GML is used to store input data and output results.
5. SVG provides a scalable visualization for the clusters.

The entire framework is presented in the following diagram (Figure 2).

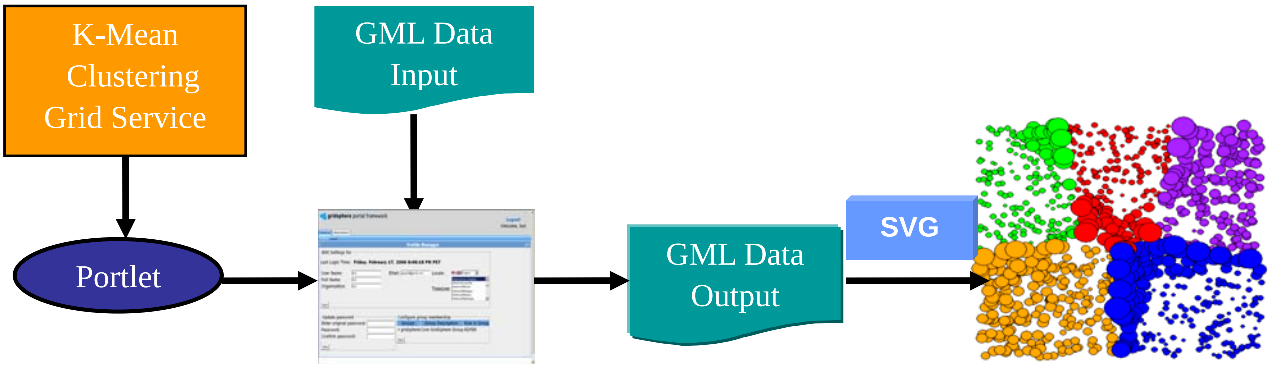

Figure 2. K-Means Clustering Demo Framework

Currently we have successfully deployed the K-Means clustering Grid service and implemented as a portlet. The results of this clustering grid service are shown in the following snapshot (Figure 3):

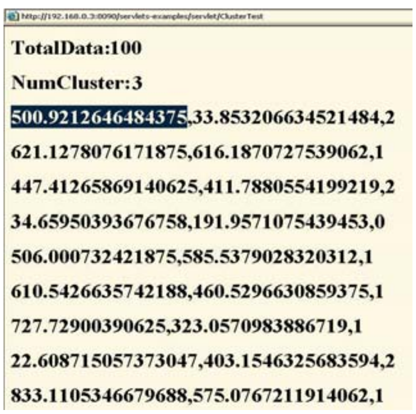

Figure 3. Results of Clustering Grid Service

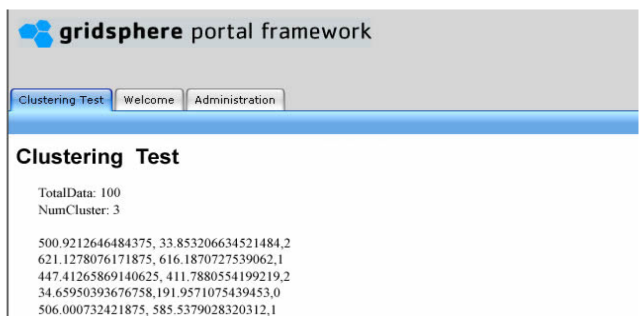

Figure 4. Clustering Portlet Running in Portlet Container

The Grid service then will be implemented in a portlet, a java-like web-based program that can be managed by portlet containers which usually are web portals. The URL of the grid service will be used as a reference in the portlet to be accessed by users. GridSphere offers a simple way to publish the portlets. The portlet (Figure 4) will generate the same clustering results as shown in figure 3.

6. Conclusions:

The GIS web portal has been providing distributed geospatial data services as well as web mapping services for a while. Current computing technologies make it possible to extend the portal to be capable of dealing with complex geospatial applications. This requires a new design of the GIS web portal. The integration of the technologies has to take into account the special concerns of geospatial problems. This paper reviews promising technologies and proposes a GIS web portal problem solving environment architecture. So far we have implemented a Grid-enabled clustering portlet in a web portal. This is a simple demo while we accumulate experiences of new techniques. The demo shows that by means of open source software, it is possible to implement analytical GIServices in a Grid environment through web portals. However, many technical challenges were encountered in the implementation. This reminds us that there will be even more problems to implementing more complex GIS analytical web portals. Major research will be expected on: extending grid computing software and Web Services frameworks to better support GIServices; developing workflow semantics for specific geospatial problems; searching and locating appropriate GIServices; selecting the best GIServices; making GIServices downloadable to improve the availability. The Further research plans will focus on choosing a real-world geospatial application and designing a GIS web portal to simulate the problem solving process. The demo design, development, and user evaluation are the major tasks processes that we expect to work on for the next two years.

Acknowledge

It is a pleasure to gratefully acknowledge the contributions made by Qinghua Qiao and Lin Xu, Wuhan University to the implementation of the demo.

References

Agarwal, P. 2005. Ontological considerations in GIScience. *International Journal of Geographical Information Science* 19(5): 501-536.

Anderson, G. and R. Moreno-Sanchez. 2003. Building web-based spatial information solutions around open specifications and open source software. *Transactions in GIS* 7(4): 447-466.

Arpinar I., A. Sheth, C. Ramakrishnan, E. Usery, M. Azami, and M. Kwan. 2005. Geospatial ontology development and semantic analytics. In: J. Wilson and A. Fotheringham (eds), *Handbook of Geographic Information Science*, Oxford: Blackwell Publishing.

Baker M., R. Buyya, and D. Laforenza. 2002. Grids and grid technologies for wide-area distributed computing. *Software - Practice and Experience* 32(15): 1437-1466.

Berners-Lee, T., J. Hendler, and O. Lassila. 2001. The Semantic Web: a new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American* 284(5): 34-43.

ESRI. 2006. GIS Web Services. http://www.esri.com/software/arcgis/about/gis-webservices.html (accessed February 25, 2006).

GEON (The Geosciences Network: Building Cyberinfrastructure for the Geosciences). 2006. http://www.geongrid.org/ (accessed February 22, 2006).

Fonseca, F., and G. Câmara. 2002. Using ontologies for integrated geographic information Systems. *Transactions in GIS* 6(3): 231-257.

Fonseca, F., C. Davis, and G. Câmara. 2003. Bridging ontologies and conceptual schemas in geographic information integration. *Geoinformatica* 7(4): 355-378.

Foster, I., and C. Kesselman. 1997. Globus: A metacomputing infrastructure toolkit. International *Journal Supercomputer Application*s 11(2):115-128.

Foster, I., C. Kesselman, and S. Tuecke. 2001. The anatomy of the grid: enabling scalable virtual organizations. *International Journal of Supercomputer Applications* 15(3): 200-222.

Globus Alliance. 2006. http://www.globus.org/ (accessed February 9, 2006).

Lin, K., and B. Ludaescher. 2004. GEON: ontology-enabled map integration in Twenty-fourth Annual ESRI International User Conference, San Diego, CA.

Mark, D.M. 2000. Geographic information science: critical issues in an emerging cross-disciplinary research domain. *URISA Journal* 12(1): 45 – 54.

Mark, D.M., B. Smith, M. Egenhofer, and S. Hirtle. 2004. Ontological foundations for geographic information science. In R.B. McMaster and L. Usery (eds), *Research challenges in geographic information science*. Boca Raton, Florida: CRC Press. pp. 335-350.

Mehmet S.A., G. Aydin, G.C. Fox, H. Gadgil, M. Pierce, and A. Sayar. 2005. Information services for Grid/Web Service Oriented Architecture (SOA) based geospatial applications. Technical Report, June, 2005. Community Grids Lab (CGL), Pervasive Technology Labs (PTL), Indiana University, Bloomington.

Miller, H. J., and J. Han. 2001. Geographic data mining and knowledge discovery: an overview. In H. J. Miller and J. Han (eds), *Geographic data mining and knowledge discovery*. London: Taylor and Francis.

Open Geospatial Consortium. 2006. Sensor web enablement and OpenGIS SensorWeb™. http://www.opengeospatial.org/functional/?page=swe (accessed February 21, 2006).

Open Source Geospatial Foundation. 2005. An open letter to the MapServer open source web mapping community. http://www.dmsolutions.ca/news/openletter-11-28-2005.html. (accessed April 8, 2006).

Pike W.A., and M. Gahegan. 2003. Constructing semantically scalable cognitive spaces. In Spatial Information Theory: Foundations of Geographic Information Science. Conference on Spatial Information Theory COSIT03, Lecture Notes in Computer Science 2825, K. Worboys M, and S. Timpf (eds), Berlin: Springer-Verlag, Berlin. pp. 332-348.

Smith, B., and D.M. Mark. 2001. Geographical categories: an ontological investigation. *International Journal of Geographical Information Science* 15(7):m591-612.

Tait, M.G. 2005. Implementing geoportals: applications of distributed GIS. Computers, *Environment and Urban Systems* 29(1): 33-47.

Tang, W., and J. Selwood. 2005. *Spatial portals: gateways to geographic information*. Redlands, California: ESRI Press.

Tsou, M.-H. 2005. The big changes in Internet GIS. *GIS@development* 9(10).

Visser, U., H. Stuckenschmidt, G. Schuster, and T. Vögele. 2002. Ontologies for geographic information processing. *Computers and Geosciences* 28(1): 103-117.

W3C (World Wide Web Consortium). 2006. Semantic Web. http://www.w3.org/2001/sw/ (accessed February 24, 2006).

Wang, S. 2006. GISolve: A grid-based problem solving environment for computationally intensive geographic information analysis. Presented at Global Grid Forum 15 GIS Workshop: Building Geographical Information System Grids.

Yang, C., D. Wong, R. Yang, M. Kafatos, and Q. Li. 2005. Performance-improving techniques in web-based GIS. *International Journal of Geographical Information Systems* 19(3): 319-342.