Transactions in GIS

Transactions in GIS, 2010, 14(6): 755–772

Research Article

GeoPW: Laying Blocks for the Geospatial Processing Web[1]

Peng Yue
State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing Wuhan University

Jianya Gong
State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing Wuhan University

Liping Di
Center for Spatial Information Science and Systems (CSISS) George Mason University

Jie Yuan
State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing Wuhan University

Lizhi Sun
State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing Wuhan University

Ziheng Sun
State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing Wuhan University

Qian Wang
State Key Laboratory of Information Engineering in Surveying Mapping and Remote Sensing Wuhan University

Abstract
Recent advances in Web-related technologies have significantly promoted the wide sharing and integrated analysis of distributed geospatial data. Geospatial applications often involve diverse sources of data and complex geoprocessing functions. Existing Web-based GIS focuses more on access to distributed geospatial data. In scientific problem solving, the ability to carry out geospatial analysis is essential to geoscientific discovery. This article presents the design and implementation of

Address for correspondence: Peng Yue, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, 129 Luoyu Road, Wuhan, 430079, China. E-mail: geopyue@gmail.com

doi: 10.1111/j.1467-9671.2010.01232.x

GeoPW, a set of services providing geoprocessing functions over the Web. The concept of the Geospatial Processing Web is discussed to address the geoprocessing demands in the emerging information infrastructure, and the role of GeoPW in establishing the Geospatial Processing Web is identified. The services in GeoPW are implemented by developing middleware that wraps legacy GIS analysis components to provide a large number of geoprocessing utilities over the Web. These services are open and accessible to the public, and they support integrated geoprocessing on the Web.

1 Introduction

The increasing use of sensors and platforms provides powerful capabilities for collecting geospatial data. For example, the four satellites of the National Aeronautics and Space Administration's (NASA's) Earth Observing System (EOS) now collect 1,000 terabytes annually (Clery and Voss 2005), far more than Earth scientists can hope to analyze. Geoscientists are experiencing a data-rich yet analysis-poor period. Recent advances in Web-related technologies have significantly promoted the wide sharing of, discovery of, and access to distributed geospatial data. As a result, many Web-based GIS applications and online geospatial data services are now available. Geospatial applications often involve diverse sources of data and complex geoprocessing functions. Existing Web-based GIS focuses more on access to distributed geospatial data. In scientific problem solving, the ability to carry out geospatial analysis is essential to geoscientific discovery. Providing geoprocessing capabilities on the Web is a critical research issue in GIS.

With the evolution of the Web and wide applications of Web Service technologies, Service-Oriented Architecture (SOA) has shown great potential for scientific research. SOA is "a way of reorganizing a portfolio of previously siloed software applications and support infrastructure into an inter-connected set of services, each accessible through standard interfaces and messaging protocols" (Papazoglou 2003). Geospatial data and geoprocessing functions are encapsulated as services with standard interfaces and protocols to allow Web-based sharing and automatic access. Foster (2005) uses the term *Service-Oriented Science* to refer to the scientific research supported by distributed networks of interoperating services. In the geospatial domain, a number of specifications for geospatial Web Services have been developed, most notably the Open Geospatial Consortium (OGC) standards-compliant services, including the Web Feature Service (WFS), Web Map Service (WMS), Web Coverage Service (WCS), Sensor Observation Service (SOS), Catalogue Services for the Web (CSW), and Web Processing Service (WPS). Among them, the OGC WPS specification is the one focusing on geoprocessing over the Web.

This article describes the design, implementation, and potential of GeoPW, a Web Service toolkit to support geoprocessing over the Web. It uses Web Services, legacy GIS software, and OGC specifications to make traditional desktop analysis functions configurable, accessible, and interoperable on the Web. The concept of the Geospatial Processing Web is discussed to address the geoprocessing demands in the emerging information infrastructure, and the role of GeoPW in establishing the Geospatial Processing Web is identified. The remainder of the article is organized as follows. Section 2 introduces related work in the literature. Section 3 introduces the background on geoprocessing over the Web and the vision of the Geospatial Processing Web. Section 4 presents the architecture design of GeoPW, and Section 5 describes the system implementation details. Conclusions and pointers to future work are given in Section 6.

2 Related Work

The growth of the Web has resulted in Web-based sharing of distributed data and computational resources. In order to enhance the collaboration, creativity, and rich user experiences in Web applications, a set of Web technologies and applications including Asynchronous JavaScript and XML (AJAX), Rich Internet Application (RIA), blog, wiki, and social networking, collectively referred to as Web 2.0 (O'Reilly 2005), are widely employed. The Semantic Web, inspired by Tim Berners-Lee (1998), inventor of the Web, provides a set of technologies to increase the intelligence of the Web. In recent years, the term Cyberinfrastructure (in the United States) or e-Infrastructure (in Europe) (Hey and Trefethen 2005) have been used increasingly to refer to a comprehensive information infrastructure, which integrates computing hardware and systems, data and information resources, networks, digitally enabled-sensors, online instruments and observatories, virtual organizations, and experimental facilities, along with an interoperable suite of software and middleware tools and services (US NSF 2007). The rapid development of Grid technologies (Foster et al. 2001) and emergence of cloud computing (Vaquero et al. 2009) are now consolidating the development of the Cyberinfrastructure.

While the existing technologies provide generic approaches for building the emerging information infrastructure, there are still substantial research challenges to develop high-level intelligent middleware services and domain-specific services for problem-solving and scientific discovery (Hey and Trefethen 2005). Scientific discovery usually requires heavy analysis and synthesis. For the geoscientific, education, and research users, current technologies do not directly address the issue of how distributed data and geoprocessing functions can be used to actually meet their geospatial analysis demands. The Geospatial Processing Web will provide a functional framework for the information infrastructure tailored to the analysis demands of geoscientific discovery.

The emergence of geoprocessing over the Web has made the WPS specification and its related applications subjects of active study in recent years (Díaz et al. 2007, Foerster and Schäffer 2007, Kiehle 2006). OGC approved version 1.0 of WPS as a standard in 2008. WPS specifies the interface and protocol for geoprocessing services. Some open source implementations of WPS such as 52n WPS (52n WPS 2006), Python Web Processing Service (PyWPS 2009), and Deegree Web Processing Service (Deegree 2009) exist. They provide frameworks and examples for implementation of geoprocessing services. Michaelis and Ames (2009) suggest some improvements in WPS implementation. The work presented in this article focuses on providing a large number of geoprocessing services as building blocks for the Geospatial Processing Web, with a greater emphasis on how to wrap legacy GIS components.

Some studies focus on integrating and chaining geoprocessing services. Stollberg and Zipf (2007) propose several approaches in WPS to support service chaining. Brauner and Schäffer (2008) demonstrate the use of the Web Services Business Process Execution Language (WSBPEL) (OASIS 2007), known as BPEL for short, in chaining WPS services. Yue et al. (2008, 2009) demonstrate the use of OWL-S (Martin et al. 2004), the Web Ontology Language (OWL) based Web Service Ontology, and BPEL for chaining WPS services. Therefore, the geoprocessing services provided by GeoPW can be chained with other standards-compliant services based on these approaches.

The Geographic Resources Analysis Support System, commonly referred to as GRASS GIS, has been used in scientific workflow systems such as the Kepler workflow system (Zhang et al. 2006). Providing geoprocessing services over the Web has been an active goal of the GeoBrain project (Di 2004). GRASS was encapsulated in GeoBrain as World Wide Web Consortium (W3C) SOAP-based Web services. Current OGC Web Services are not equivalent to the W3C SOAP-based Services (Tu and Abdelguerfi 2006). Most OGC Web Service implementations provide access via HTTP GET and HTTP POST, and they do not support SOAP. Providing WPS services through GRASS is a promising approach. PyWPS and 52n WPS (Brauner and Schäffer 2008) demonstrate the WPS interface for GRASS. From this aspect, the work presented in this article is similar to theirs. In particular, the work provides a flexible way to develop processes initiated by commands and provides processes covering many of the GRASS analysis functions instead of only several examples.

3 Geoprocessing Over the Web

The term geoprocessing has a broader meaning than the analysis in traditional GIS and can refer to any sort of geospatial processing or analysis function. For example, it includes computational models such as Earth system prediction models, which are not available in traditional GIS. In real world applications, a single analysis function is not enough to solve a complex problem. Therefore, geoprocessing models, which combine multiple analysis functions to solve application-specific problems, are needed. Environmental Systems Research Institute (Esri) provided a geoprocessing framework in its ArcGIS 9 software product released in 2004. This framework includes a number of analysis functions (sometimes called processes). In addition to these basic functional analysis units, a ModelBuilder tool is provided to help users visually combine different analysis processes and generate geoprocessing models (or called process chains). The spatial modeling tool in the ERDAS IMAGINE software has a similar process chaining function. It can combine GIS analysis functions with ERDAS's commercial image processing functions. Other traditional GIS software systems support common analysis functions such as overlay, buffer, and network analysis. Nevertheless, these legacy GIS software systems cannot support geoprocessing over the Web. The analysis functions and even the process modeling capability can be used only in their own proprietary environments.

From the viewpoint of SOA, geoprocessing over the Web should follow the publish-find-bind paradigm. The individual geoprocessing services are published in a service registry. The geoprocessing requestor is able to find geoprocessing services through the registry. When the appropriate services are located, the requestor can bind and assign geoprocessing tasks. In a distributed environment, a complex geoprocessing task may require services provided by multiple service providers. A geoprocessing model that contains multiple geoprocessing steps will be bound to an executable service chain or concrete workflow.

The Geospatial Processing Web named here refers to a distributed, integrated, and collaborational service-oriented geoscientific research environment that geoprocesses over the Web. It is an information infrastructure where various geoprocessing functions

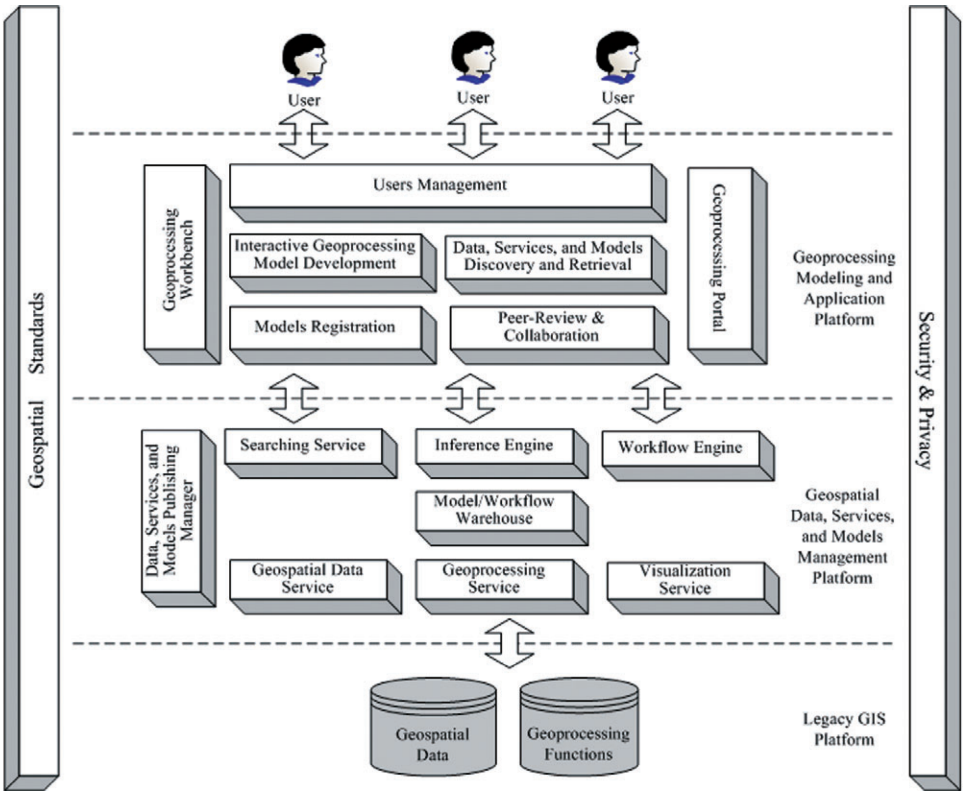

Figure 1 Geospatial processing web framework

are exposed as services and accessed through standard protocols. These services can be located and integrated intelligently for complex geoprocessing tasks. The Web is evolvable by accommodating new geoprocessing components and models, and maintainable through upgrade or replacement of existing geoprocessing components and models.

Figure 1 illustrates the framework of the Geospatial Processing Web and components to support it. It is a three-layer architecture. The first layer is the legacy GIS platform that manages geospatial data and analysis functions in its own proprietary form. The second layer is the geospatial data, services, and models management platform, which provides basic utilities for geoprocessing modeling and applications in the higher layer. Such basic utilities include services to provide, process, and visualize geospatial data. For process chaining, geoprocessing models and workflow management utilities are needed. The search service can be used to locate geoprocessing services or models, and inference engines may be required for a semantic match when needed. In the third layer, a workbench or portal can be used to help users develop geoprocessing models and applications. The geoprocessing models, after going through a collaborative peer review, can be registered in the model warehouse as a type of knowledge.

Implementation of the Geospatial Processing Web combines the latest developments in distributed computing technologies, Web technologies, and interoperability standards. Grid technologies allow sharing of distributed geospatial data and computational resources and coordinated problem solving among dynamic virtual organizations. Cloud computing provides powerful information technology (IT) services. The Web 2.0 technologies bring rich user experience in geoscientific analysis, and Semantic Web approaches increase the intelligence of information discovery. Geospatial standards ensure the interoperability of systems for the Geospatial Processing Web. Using the standards for Web Services, multiple services can be discovered and automatically integrated on the Web. This integration result can be accessible as a new service. Service-oriented applications can increase individual and collective scientific productivity by making powerful information tools available to scientists, allowing widespread automation of data analysis and computation (Foster 2005). The standards-based interoperable architecture adopted in the implementation of the Geospatial Processing Web allows the plug-and-play of community-developed, standards-compliant Web services. Development of standards-compliant, interoperable, distributed service components will ensure the openness, growth, and evolution of the Geospatial Processing Web. Since GeoPW provides a set of geoprocessing services, it serves as the building block for higher geoprocessing modeling and complex geoprocessing tasks, and it is part of the fundamental utilities for the implementation of the Geospatial Processing Web.

4 GeoPW: System Design

Figure 2 shows the system architecture of GeoPW. GeoPW not only supports the provision of a warehouse for geoprocessing service modules, but also enables development of new geoprocessing services to be easily carried out. It provides this warehouse by wrapping legacy or new geoprocessing functions. The following properties must be considered when developing geoprocessing services.

4.1 Granularity of Geoprocessing Services

The warehouse contains individual geoprocessing Web Services that are self-contained, self-described, and modular. The service modules can be published, located, and dynamically invoked or chained together on demand to form complex geoprocessing models for producing sophisticated geospatial information products. The granularity of individual service modules is an important factor affecting the flexibility, applicability, and reusability of service modules in different geospatial models. If a module's functionality is too small, many modules are needed to construct a complex geospatial model, hence reducing the system performance. If too many functions are aggregated into a service module, the module will not be easily plugged into other geospatial models. The flexibility, applicability, and reusability will decrease.

The International Organization for Standardization (ISO) 19119 Geographic Information – Services standard (ISO/TC 211 2005), defines six categories of geospatial services: geographic human interaction services, geographic model/information management services, geographic workflow/task management services, geographic processing services, geographic communication services, and geographic system management services. The geographic processing services can be further divided into four categories by whether they modify the following aspects of geospatial data: spatial, thematic, temporal, and metadata. Most geoprocessing services can be classified into these four categories. However, ISO 19119 is a high-level service architecture standard, and it provides no fine-grained geoprocessing service taxonomy. In the science keywords provided by the

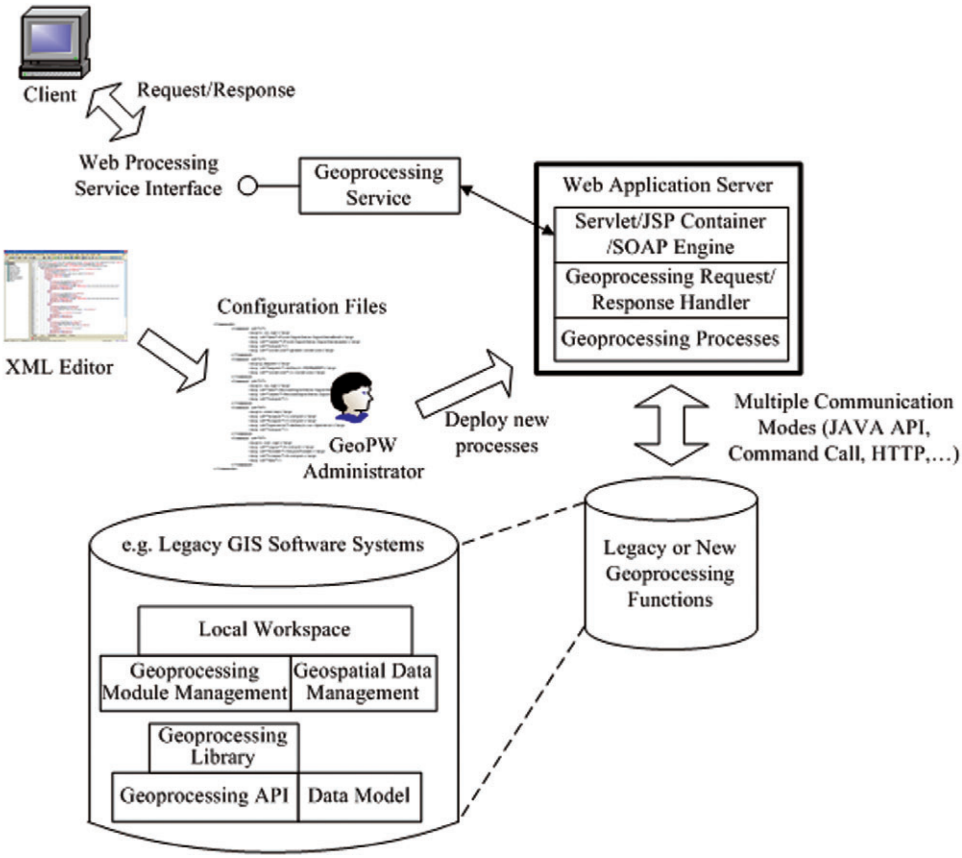

Figure 2 GeoPW system architecture

Global Change Master Directory (GCMD) (Olsen et al. 2007), a category of service keywords for Earth science is defined hierarchically. GIS functionality is classified as a sub-category in the Data Analysis and Visualization topic. The commercial GIS software system ArcGIS provides over 200 geoprocessing tools. These geoprocessing tools (ESRI 2006) are grouped into over twenty toolboxes, each representing a kind of geoprocessing function. The geoprocessing functions in GRASS are grouped into vector and raster analysis categories (GRASS 2008). Over one hundred geoprocessing commands are available in these two categories. The geoprocessing tools in ArcGIS and geoprocessing commands in GRASS are typical examples of the functional granularity in GIS. Therefore, these fundamental geoprocessing functions in GIS software systems, designed originally to be tailored to users' preferences, can provide a valuable reference for determining the granularity of geoprocessing services.

4.2 Reuse of Existing Components

The service modules in GeoPW act as individual building blocks for dynamically constructing complex geoprocessing models. The richness of the service modules will, to a certain extent, determine the capability of the in-house processing power of the GeoPW system. Through many years of development, existing GIS software systems have accumulated various analysis functions, and they can be reused. Therefore, the key development effort for GeoPW resides in the middleware service layer, focusing on developing various geoprocessing processes that communicate with legacy analysis components through a range of modes including JAVA Application Programming Interface (API) call, Command call, or HTTP call. As shown in Figure 2, the Web application server hosts services with various geoprocessing processes that will be accessed by different clients. The request/response messages for geoprocessing services can be processed initially by either the servlet/JSP container or the SOAP engine. The Simple Object Access Protocol (SOAP) is a Web Service standard established by W3C (W3C 2001). The payload of the message, i.e. the content specific for geoprocessing services, is further processed by service request/response handlers. The corresponding geoprocessing process invokes a geoprocessing task. For example, if geoprocessing services are implemented using JAVA, there will be JAVA classes to implement request/response handlers. And geoprocessing operations will be implemented as JAVA methods.

The existing geoprocessing component in a legacy GIS software system is a suite of software packages sharing common runtime features (Figure 2). For example, it has the same geoprocessing module and data management, API style for geoprocessing and data operation, for example Dynamic-Link Library (DLL)-based API or Command-based API, and environment variables settings. Therefore, once a geoprocessing process wrapping a certain analysis function of a software package is implemented, it provides a skeleton for other geoprocessing processes wrapping the same software package. To avoid the duplication of the same steps for implementing various geoprocessing processes wrapping the same software package, the design pattern template method can be used (Gamma et al. 1995). The skeleton for the targeted legacy analysis component is defined, deferring some specific steps to subclasses or configuration files. As is shown in Figure 2, GeoPW administrators can create, edit, and deploy their own XML-based configuration files to develop new geoprocessing processes, with little or no knowledge of changing, compiling, or building computer source code.

4.3 Adherence to Standards

The standards implemented in GeoPW make the system interoperable, reusable, extensible, and evolvable. In an interoperable service environment, service modules existing in different provider communities can and will be shared. Therefore, a complex geospatial model can be built from service modules in all member systems of the provider communities and can be executed across the member systems. This is the power of standards and federation. The GeoPW service modules are developed to follow the OGC WPS standard that is currently under development.

OGC WPS specifies a standard interface and protocol for discovering and executing distributed geoprocessing processes. A geoprocessing process could be any kind of GIS analysis function. The three mandatory operations included in the standard WPS interface are GetCapabilities, DescribeProcess, and Execute (Schut 2007). The GetCapabilities operation allows a client to request and receive a service capabilities document that describes the operations and processes of a specific WPS implementation. The DescribeProcess operation allows a client to get detailed information, such as input and output parameter types, about specific processes. The Execute operation allows the client to run a specific process in a WPS server. In addition to the service interface, more specific features are defined: encodings of request/response for process execution, embedded data and metadata in inputs/outputs for process execution, references to Web-accessible data inputs/outputs, long-running processes support, process status information and processing errors report, and storage of process outputs. More detail is provided in the OGC WPS document (Schut 2007). Like other OGC service specifications, WPS uses the HTTP protocol, in particular the GET and POST operations, to exchange XML-based request and response messages.

5 Implementation

The 52n WPS framework (52n WPS 2006) is used to develop all geospatial processing services in GeoPW. It runs as a Web application in the Jakarta Tomcat server. The analysis functions provided by the GRASS and GeoStar GIS software systems are wrapped as geoprocessing processes and plugged into the 52n WPS.

5.1 Wrapping GRASS and GeoStar by using the 52n WPS Framework

GRASS is one of the most widely used open source GIS software systems. Originally developed by the US Army Construction Engineering Research Laboratories (USA-CERL), GRASS has evolved into a powerful utility with over 350 programs and tools ranging from vector/raster analysis to data visualization and image processing. GRASS is integrated with the Geospatial Data Abstraction Library (GDAL/OGR) (GDAL 2008) to support an extensive range of raster and vector formats, including the OGC-conformant Simple Features (GRASS 2008). A distinguishing feature of GRASS is that it provides command line syntax for its large geoprocessing functions.

GeoStar is a leading enterprise GIS software system in China. It was developed, starting in 1992, by the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS) of Wuhan University (formerly Wuhan Technical University of Surveying and Mapping) (Li et al. 1999). The software is a suite of functional components for feature and image data management, processing, and visualization. The spatial analysis component in the current version of GeoStar is a DLL with C++ API running under Windows.

The primary problem in implementation is using the legacy GIS analysis components with the OGC WPS interface and protocol. The 52n WPS framework is a JAVA Web application. It provides servlets and general request and response handlers to process HTTP GET/POST requests for WPS GetCapabilities, DescribeProcess, and Execute operations. WPS is a generic interface and does not specify any particular processes. Therefore, the implementation of WPS must define the processes that it supports. The 52n WPS framework provides a pluggable framework to allow new processes to be added and easily accessed through the OGC WPS interface and protocol. As shown in Figure 3, the framework defines two kinds of repositories. The first one is the Algorithms Repository. An abstract JAVA class AbstractAlgorithm is defined and must be inherited by all implementation classes for the corresponding geoprocessing processes. These processes are registered in the algorithms repository through a configure file and loaded into the server during the initialization of a WPS server. Each process has its XML-based process description document that can be retrieved by the DescribeProcess operation. The second repository is the Data Handler Repository. The input and output parameters

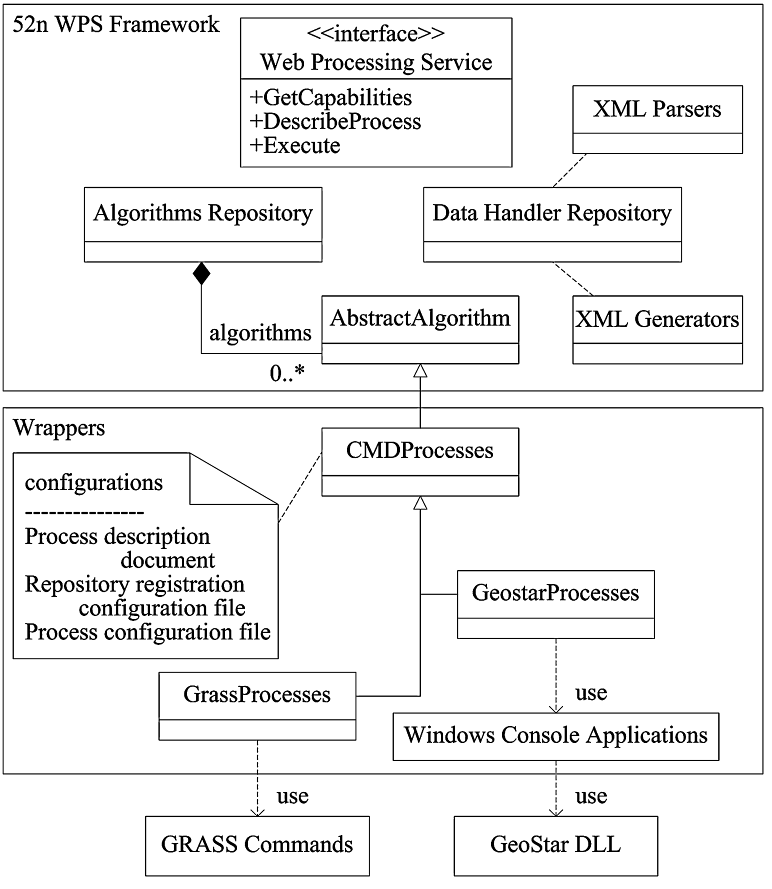

Figure 3 Wrappers for GRASS and GeoStar in the 52n WPS framework

for specific processes may be complex or simple. A simple parameter may be described by a simple string or integer, while a complex parameter needs an XML schema to describe its structure (e.g. a feature collection represented using GML). To generate or parse complex parameter values, the XML generator and parser classes must be registered in the data handler repository and loaded during the initialization of the WPS server.

Figure 3 illustrates the development of wrappers for GRASS and GeoStar, so that legacy analysis functions can be wrapped as new processes in the 52n WPS framework. As mentioned in Section 4.2, the design pattern Template Method is used. An abstract class CMDProcesses is developed to invoke commands and communicate with their execution to log status. GRASS offers commands and shell scripts that can be invoked through methods in this class. The use of GRASS commands requires the establishment of a working environment including DATABASE, LOCATION, and MAPSET. Therefore, a sub-class GrassProcesses inherited from CMDProcesses is defined to handle these specific details. Similarly, a class GeostarProcesses is inherited from CMDProcesses. As mentioned before, the spatial analysis component of GeoStar is a Windows DLL. Therefore, a Windows Console Application is provided to wrap the DLL functions as

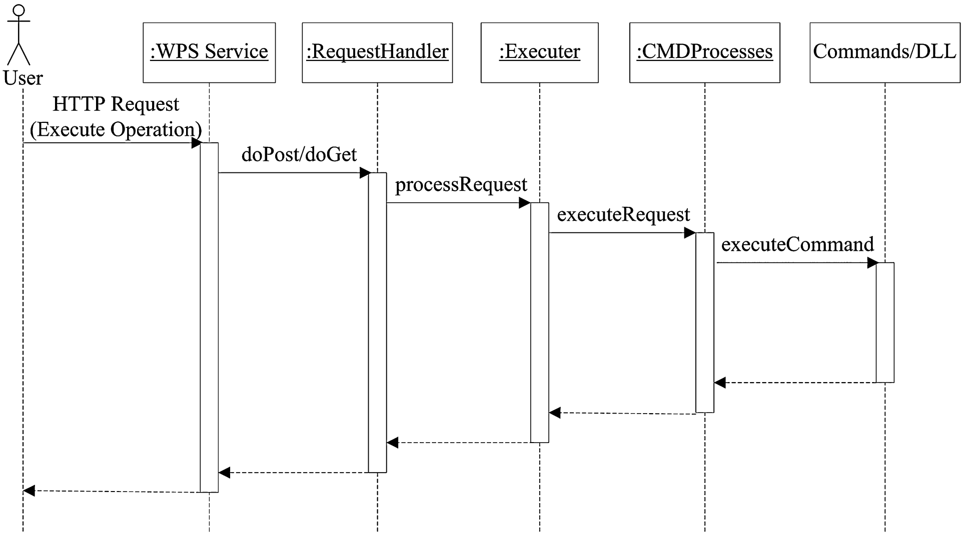

Figure 4 Sequence diagram of requests execution

commands. The sequence that the geoprocessing service uses to process a WPS Execute operation is illustrated in Figure 4.

The wrappers provide a way to develop new processes using XML scripting. An XML-based process configuration file describes the sequence of internal command steps that are needed to complete a geoprocessing function defined by a process. For example, a single process using GRASS requires the following steps: importing, processing, and exporting data. In addition, the process configuration files describe the data bindings between the command parameters and input/output of a process description. Therefore, the wrappers can generate executable commands or scripts based on these configuration files and deliver the output to the WPS response. These process configuration files, together with the process description documents and registration configuration file for repositories, define the processes developed and deployed. The development of new processes can then be simplified to adding new XML configuration files to compose and expose analysis functions. If a bug is detected or a new version of software or an algorithm is available, developers can update the configuration files or GIS software without changing the service interfaces, thus eliminating the manipulation of original computer code and software upgrades by service end users.

5.2 Services Provided

GeoPW has already developed many geoprocessing services using the functions provided by GRASS and GeoStar. These services provide geoprocessing functions including GIS vector and raster analysis, network analysis, and data processing. They are grouped into ten categories: Vector Spatial Analysis, Vector Geometry Operation, Vector Network Analysis, Vector Data Processing, Raster Spatial Analysis, Raster Data Processing, Raster Aggregation, Hydrologic Analysis, Geostatistics, and Others. Figure 5 shows the number of services in different categories.

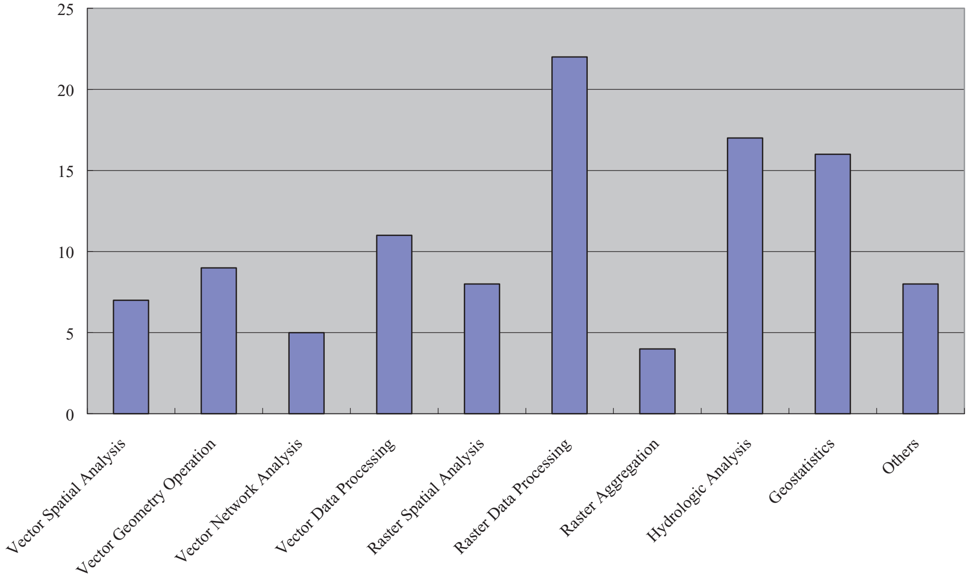

Figure 5 Distribution of services in different categories of geoprocessing functions

All these services are deployed on the Hewlett-Packard (HP) blade servers (Intel® Xeon® 5110 1.6GHz, 4GB RAM) located at LIESMARS. GRASS software is installed in a server blade with a UNIX operating system, and GeoStar is located in another server blade with the Microsoft Windows operating system. All services are provided to the public through a single Web access point on one server blade that is publicly accessible. This Web entry point is implemented using servlets to redirect the request to processes deployed on internal server blades. Currently, the services in GeoPW have wrapped most vector and raster analysis functions of GRASS and demonstrated the GeoStar wrapper by exposing its spatial buffer and overlay analysis functions. Additional processes are being developed, and after passing tests, they will be deployed into distributed internal server blades.

The latest news about GeoPW and all developed geoprocessing services is available at http://geopw.whu.edu.cn/geopw.html (Figure 6). The process description and one testing example for each process are provided through Web page links. A link on the home page provides a guide for GeoPW. This guide provides Web addresses for geoprocessing services and process descriptions and testing examples for analysis functions provided by both GeoStar and GRASS. Information on how to invoke these services through the HTML Form and compose services using Oracle BPEL software is also available on the guide page.

The services developed in GeoPW are being integrated with other geospatial technologies to support additional functional aspects of the Geospatial Processing Web. For example, the services in GeoPW are now integrated with Virtual Globe software. The emergence of Virtual Globe software systems (Declan 2006) such as Google Earth, Microsoft Virtual Earth, and NASA World Wind provides a new generation of learning tools to help students digest large-scale geospatial information about the world, and supports expert analyses of domains in an interactive three-dimensional virtual environ-

Figure 6 Web access entry for GeoPW

ment. The integration is tested through the GeoGlobe system, a leading enterprise Virtual Globe software system in China. Figure 7 shows the result of computing terrain slope in the GeoGlobe. The geoprocessing functions are listed on the left panel of the GeoGlobe client. Such integration combines the visualization and communication power of Virtual Globes with the powerful analysis functionalities of geoprocessing services, thus providing an analysis-enhanced extensible Virtual Globe to help students and researchers investigate various scientific problems in an environment with natural and intuitive user experiences.

Another example is the development of the geoprocessing modeling tool. An online geoprocessing model designer, called GeoPWDesigner, has been developed using Asynchronous JavaScript and XML (AJAX) technology. The design of the tool follows the three-phase procedure for service composition (Yue et al. 2009): (1) process modeling, which generates a composite process model consisting of the control flow and data flow among atomic processes; (2) process model instantiation, where the abstract process is instantiated into a concrete service chain; and (3) workflow execution, where the chaining result is executed in the workflow engine to generate the desired data product. Figure 8 illustrates an example of a geoprocessing model for a flooding analysis. Atomic process models on the left column of the palette can be dragged manually to the right panel. Different process models can be composed using data flow and control flow links to generate a composite process model. The process model can be transformed into a service chain composed of GeoPW services or other interoperable services. The chaining result is executed by a BPEL workflow engine.

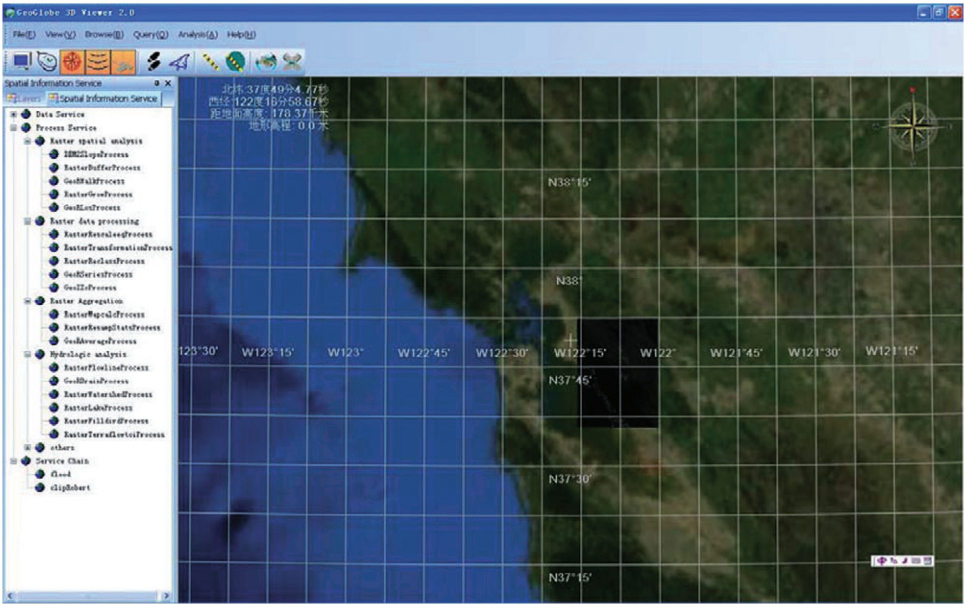

Figure 7 Integration of GeoPW and GeoGlobe

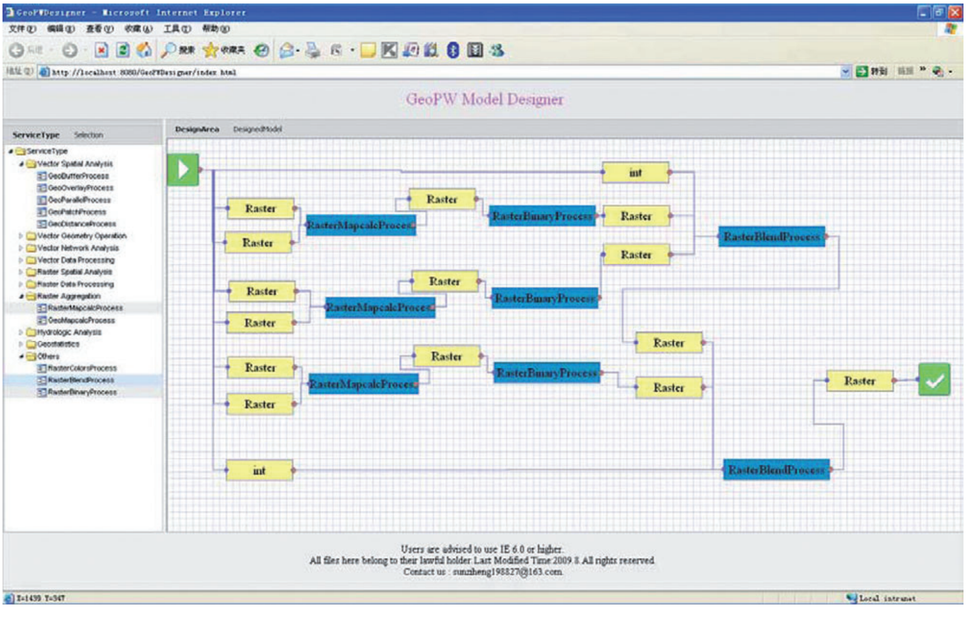

Figure 8 GeoPWDesigner interface, showing geoprocessing modeling capability

5.3 Results Analysis

The effectiveness of the approach is demonstrated by the fact that over one hundred geoprocessing services have been developed successfully for GRASS and GeoStar. Both vector and raster analysis services have been developed. In addition, customized geoprocessing services through combining commands in GIS can also be provided. Although GML has been widely used as the geospatial data encoding format over the Web, popular GIS file formats such as ESRI Shapefile are supported, since these file formats are commonly used for geospatial data owned by users and therefore, practical and ready-to-use. These interoperable services in GeoPW can be plugged into the Virtual Globe software or a geoprocessing modeling tool. As also demonstrated by Yue et al. (2008, 2009), standards-based services can be semantically described, published, discovered, and chained to support the Geospatial Processing Web mentioned in Section 3.

Through the research and implementation experiments, several issues that are important for the practical use of geoprocessing services have been identified:

1. Process with a long running time: A geoprocessing process may take a long time to execute, especially when such a processing is data-intensive or computing-intensive. For data-intensive processing, although a FTP server can be installed for the purpose of temporary dataset uploading and downloading, it is better to deploy GeoPW in a data-rich environment, possibly data pools in data archiving centers, thus upgrading data archiving centers to data-rich learning and research centers for remote end users. For computing-intensive processing, asynchronous notification should be supported by WPS implementations.
2. Standardization of specific process descriptions: As mentioned before, WPS is a generic interface and does not specify the process descriptions of specific processes. The process descriptions for various geoprocessing functions are different, such as the selection of vector, raster, text input data, or multiple output data files for a given process. These process descriptions should be regulated and presented semantically. For example, consider the buffer and overlay processes in GeoPW. Services based on GRASS and GeoStar can share the same process descriptions. Semantic Web approaches or WPS profiles (Schut 2007) provide prospects for a solution of this issue.
3. Intellectual property: Creators should be credited for their contributions of geospatial data products, algorithms, models, and services. These contributions must be peer reviewed, rated, or honored in the open Web environment according to some set of rules. Intellectual property rights should be addressed so that individuals or groups have an incentive to make contributions and feel comfortable in doing so (Foster 2005). Only in this way will more persons or organizations volunteer to contribute geoprocessing services to the Web. This community-involved, open, and cumulative approach will help drive the geoscience research environment into a cyber research service network.

6 Conclusions and Future Work

This article presents the vision and functional aspects of the Geospatial Processing Web, which has been created because of the analysis demands for the geoscientific discovery in the emerging information infrastructure. As a first step towards the establishment of the Geospatial Processing Web, GeoPW, a geoprocessing service toolkit, has been designed and implemented. The services are provided by middleware that wraps legacy GIS analysis components to provide the OGC WPS interface. These legacy components include the analysis utilities of the GIS software systems GRASS and GeoStar. Such an implementation allows the traditional desktop analysis functions to be configurable, accessible, and interoperable on the Web. Over one hundred geoprocessing services are now available in GeoPW for geoscientific research. More are being added. Among the geoprocessing functions provided by these services are GIS vector and raster analysis, network analysis, data processing, geostatistical analysis, and hydrological analysis. These services provide building blocks and a solid technological foundation for the Geospatial Processing Web.

The services provided by GeoPW are being combined with other geospatial technologies to support integrated geoprocessing on the Web. As mentioned in Section 5.2, some work has been conducted on the integration of GeoPW and Virtual Globes, and the development of geoprocessing modeling tools to support complex geospatial analysis. Further investigations will include the investigation of how GeoPW can be integrated into the existing grid or cloud computing architecture. The experience in developing WPS processes also helps guide the semantic description, discovery, and chaining of geoprocessing services. One potential is to incorporate previous work on geospatial semantic web services (Yue et al. 2008, 2009) into GeoPW to support service discovery, chaining, and geoprocessing modeling, thus enabling GeoPW to be involved into a system embracing full aspects of the Geospatial Processing Web.

Acknowledgements

We are grateful to the anonymous reviewers and Dr Barry Schlesinger for their comments. This work was funded jointly by National Basic Research Program of China (2011CB707105), Project 40801153 and 40721001 supported by NSFC, 2009 ISPRS Scientific Initiative Project, 863 Program of China (2007AA120501, 2007AA12Z214), LIESMARS and SKLSE (Wuhan University) Special Research Funding.

Endnote

1 This was one of the three best paper awards at the Ninth International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2009), Maynooth, Ireland, 7 & 8 December 2009.

References

52n WPS 2006 52n Web Processing Service (WPS). WWW document, https://www.incubator52n.de/twiki/bin/view/Processing/52nWebProcessingService

Berners-Lee T 1998 Semantic web road map. WWW document, http://www.w3.org/DesignIssues/Semantic.html

Brauner J and Schäffer B 2008 Integration of GRASS functionality in web based SDI service chains. In *Proceedings of the 2008 Free and Open Source Software for Geospatial Conference (FOSS4G 2008)*, Cape Town, South Africa: 420–9

Clery D and Voss D 2005 All for one and one for all. *Science* 308: 809

Declan B 2006 Virtual globes: The web-wide world. *Nature* 439: 776–8

Deegree 2009 Deegree web processing service. WWW document, http://www.deegree.org/

Di L 2004 GeoBrain: A web services based geospatial knowledge building system. In *Proceedings of the 2004 NASA Earth Science Technology Conference*, Palo Alto, California: 1–8

Díaz L, Costa S, Granell C, and Gould M 2007 Migrating geoprocessing routines to web services for water resource management applications. In *Proceedings of the Tenth AGILE Conference on Geographic Information Science*, Aalborg, Denmark: 1–9

ESRI 2006 An overview of commonly used tools. WWW document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=An_overview_of_commonly_used_tools

Foerster T and Schäffer B 2007 A client for distributed geoprocessing on the Web. In Ware J M and Taylor G E (eds) *Web and Wireless Geographical Information Systems: Proceedings of the Seventh International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2007), Cardiff, UK*. Berlin, Springer Lecture Notes in Computer Science Vol. 4857: 252–63

Foster I 2005 Service-oriented science. *Science* 308: 814–17

Foster I, Kesselman C, and Tuecke S 2001 The anatomy of the grid: Enabling scalable virtual organizations. *International Journal Supercomputer Applications* 15: 200–22

Gamma E, Helm R, Johnson R, and Vlissides J 1995 *Design Patterns: Elements of Reusable Object-Oriented Software*. Harlow, Addison-Wesley Longman

GDAL 2008 GDAL – Geospatial Data Abstraction Library. WWW document, http://www.gdal.org/

GRASS 2008 Geographic Resources Analysis Support System (GRASS). WWW document, http://grass.itc.it/

Hey T and Trefethen A E 2005 Cyberinfrastructure for e-science. *Science* 308: 817–21

ISO/TC 211 2005 *Iso19119:2005, Geographic Information – Services*. Geneva, International Standards Organization

Kiehle C 2006 Business logic for geoprocessing of distributed geodata. *Computers and Geosciences* 32: 1746–57

Li D, Gong J, Zhu Q, and Zhu X 1999 GeoStar: A China made GIS software for digital earth. In *Proceedings of the International Symposium on Digital Earth*, Beijing, China: 483–88

Martin D, Burstein M, Hobbs J, Lassila O, McDermott D, McIlraith S, Narayanan S, Paolucci M, Parsia B, Payne T, Sirin E, Srinivasan N, and Sycara K 2004 OWL-based Web service ontology (OWL-S). WWW document, http://www.daml.org/services/owl-s/1.1/overview/

Michaelis C D and Ames D P 2009 Evaluation and implementation of the OGC web processing service for use in client-side GIS. *GeoInformatica* 13: 109–20

O'Reilly T 2005 What is Web 2.0. WWW document, http://oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

OASIS 2007 Web services business process execution language, version 2.0. WWW document, http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf

Olsen L M, Major G, Shein K, Scialdone J, Vogel R, Leicester S, Weir H, Ritz S, Stevens T, Meaux M, Solomon C, Bilodeau R, Holland M, Northcutt T, and Restrepo R A 2007 NASA/Global Change Master Directory (GCMD) earth science keywords (version 6.0.0.0.0). WWW document, http://gcmd.nasa.gov/Resources/valids/archives/keyword_list.html

Papazoglou M P 2003 Service-oriented computing: concepts, characteristics and directions. In *Proceedings of the Fourth International Conference on Web Information Systems Engineering (WISE 2003)*, Roma, Italy: 3–12

PyWPS 2009 Python web processing service. WWW document, http://pywps.wald.intevation.org/

Schut P 2007 OpenGIS® web processing service, version 1.0.0. Wayland, MA, Open Geospatial Consortium Technical Report No. 05-007r7

Stollberg B and Zipf A 2007 OGC web processing service interface for web service orchestration: Aggregating geoprocessing services in a bomb threat scenario. In Ware J M and Taylor G E (eds) *Web and Wireless Geographical Information Systems: Proceedings of the Seventh International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2007), Cardiff, UK*. Berlin, Springer Lecture Notes in Computer Science Vol. 4857: 239–51

Tu S and Abdelguerfi M 2006 Web services for geographic information systems. *IEEE Internet Computing* 10: 13–5

US NSF 2007 *Cyberinfrastructure Vision for 21st Century Discovery*. Washington, DC, US National Science Foundation

Vaquero L M, Rodero-Merino L, Caceres J, and Lindner M 2009 A break in the clouds: Towards a cloud definition. *SIGCOMM Computer Communication Review* 39: 50–5

W3C 2001 Simple Object Access Protocol (SOAP), version 1.2. WWW document, http://www.w3.org/TR/2001/WD-soap12-part0-20011217/

Yue P, Gong J, and Di L 2008 Automatic transformation from semantic description to syntactic specification for geoprocessing service chains. In Bertolotto M, Ray C, and Li X (eds) *Web and*

Wireless Geographical Information Systems: Proceedings of the Eighth International Symposium on Web and Wireless Geographical Information Systems (W2GIS 2008), Shanghai, China. Berlin, Springer Lecture Notes in Computer Science Vol. 5373: 50–62

Yue P, Di L, Yang W, Yu G, Zhao P, and Gong J 2009 Semantic web services based process planning for earth science applications. *International Journal of Geographical Information Science* 23: 1139–63

Zhang J, Pennington D D, and Michener W K 2006 Automatic transformation from geospatial conceptual workflow to executable workflow using GRASS GIS command line modules in Kepler. In Alexandrov V N, van Albada G D, Sloot P M A, and Dongarra J (eds) *Computational Science: Proceedings of the Sixth International Conference. Berlin*. Berlin, Springer Lecture Notes in Computer Science Vol. 3993: 912–19