Web Service-based GeoBrain Online Analysis System (GeOnAS)

Liping Di, Peisheng Zhao, Weiguo Han, Yaxing Wei, Xiaoyan Li
Center for Spatial Information Science and Systems (CSISS), George Mason University
6301 Ivy Lane, Suite 620
Greenbelt, MD 20770
ldi@gmu.edu

***Abstract* – Recent advances in Service-Oriented Architecture (SOA) and in geospatial Web services and interoperability technologies are shifting geospatial data and analysis from the everything-locally-owned-and-operated paradigm to the everything-shared-over-the-Web paradigm (Web-and-Service-Centric paradigm). By embracing geospatial content and capabilities within the context of SOA, we have developed a Web Service-based Online Analysis System (GeOnAS), a fully extensible online analysis system for using GeoBrain Web services to discover, retrieve, analyze, and visualize geospatial and other network data. The most distinctive characteristic of GeOnAS is that it is composed of a set of plug-in Web services to implement all required functions. GeOnAS provides value by its overall efficiency and accuracy in integrating and analyzing distributed heterogeneous geospatial data over the Web.**

I. INTRODUCTION

Earth System Science (ESS) research and applications often involve collection, analysis, and visualization of distributed heterogeneous geospatial data. Traditionally, scientific analysis of geospatial data takes place at scientists' local machines with local data collection and software. In order for such a scenario to work, scientists must buy the computer and storage devices, spend considerable time installing and learning to use a variety of software on local machines, obtain the data from various data sources, preprocess the data locally for ingesting the data into the analysis package, and analyze the data by using the analysis packages at the local machines. This paradigm, everything locally owned and operated, makes analysis and application of geospatial data very expensive. Recent advances in Service-Oriented Architecture (SOA) and geospatial web services and interoperability technologies are shifting geospatial data and analysis from this locally oriented paradigm to sharing everything over the Web (Web-and-Service-Centric paradigm)

A significant number of geospatial datasets are now available over the Web. Many are served through standard data access interfaces, such as the Web Coverage Services (WCS) [1] of the Open Geospatial Consortium (OGC). A number of Web-based geospatial processing services have been developed and deployed over the Web. These services perform various geospatial processing functions. It is not an exaggeration to say that in the near future almost all data and analysis functions scientists need for their research will be available on the Web. The question is how to make those data and analysis service resources easily accessible to and usable by the user community. Given the above background, we have developed a Web Service-based Online Analysis System (GeOnAS), a fully extensible online analysis system for using GeoBrain Web services to discover, retrieve, analyze, and visualize geospatial and other network data. The most distinctive characteristic of GeOnAS is its composition from a set of plug-in Web services

1) Online-- users only need to have an Internet-connected PC with a Web browser;
2) OGC standards-based data access-- the system provides a single point of entry to the geospatial data from any OGC compliant data services in the world;
3) Web service-based analysis-- all functions are provided through interoperable Web services, allowing users of the system to easily perform a specific analysis by integrating new services or chaining built-in services.

The reminder of this paper is organized as follows. In section 2, the GeOnAS architecture and components are described. Section 3 illustrates how to publish and discover geospatial data and services by using catalog services. Section 4 discusses how to integrate the wide varieties of geospatial processing services to analyze geospatial data. Finally, section 5 presents the conclusions and plans for future work.

II. GeOnAS ARCHITECTURE

Local constraints SOA uses loosely coupled and interoperable Web services to implement system requirements. All the services within the SOA are independent, so that they can be accessed in a standard way without the user needing to know how the service actually performs its tasks. Moreover, SOA can support integration and orchestration of fine-grained services into coarse-grained composite services. The GeOnAS architecture shown in Figure 1 relies on SOA as its fundamental design principle for linking distributed computational resources to support geospatial analysis. The entire architecture is developed using OGC and ISO Technical Committee 211 – Geographic Information / Geomatics (ISO/TC 211) standards.

The client is a standard Web browser like Mozilla, Firefox, or Microsoft Internet Explorer. Users should not have to download and install any new software. DHTML and JavaScript techniques are used to manipulate data objects in the browser. By using AJAX (Asynchronous JavaScript and XML), the client becomes more responsive and interactive to improve the user experience.

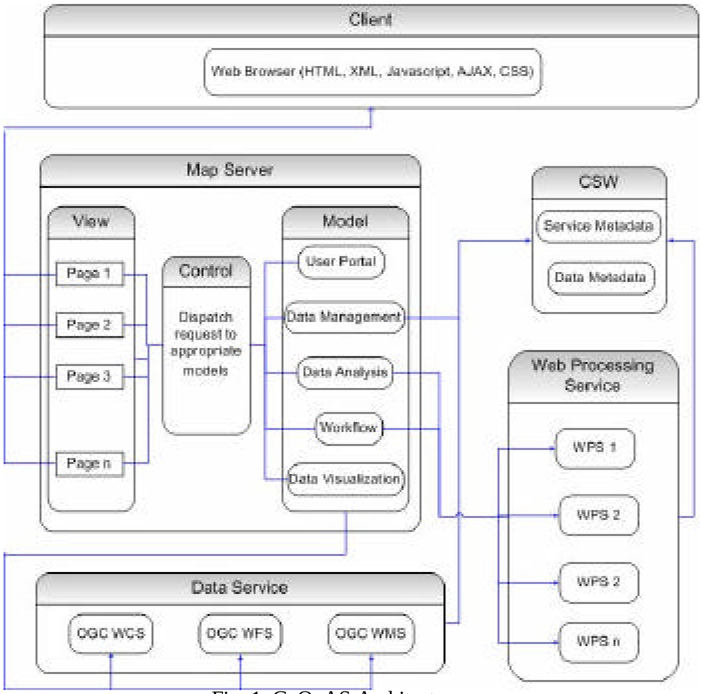

Fig. 1. GeOnAS Architecture

The map server follows the Model-View-Controller (MVC) design pattern. Typically, models represent the information of the application and act as functions by providing "get" methods to retrieve particular pieces of information or state and "set" methods to update the information and state of the model. When "set" methods are invoked, views are notified to make some changes by calling listeners that are registered with the model. Map server includes following models:

- **User Portal**: Each user is able to store the current state of the portal in an OGC Web Map Context (WMC) document. This document can be imported again later to restore the portal to its earlier state. The WMC allows users to build their own maps and systems by including a list of all preferred data layers and relevant processing services.
- **Data Management**: Each user can retrieve geospatial data using a remote service and store them on the map server temporarily, with a network accessible point. All meta-information of the saved data is encoded in XML for further use.
- **Data Analysis**: Each user is able to select or integrate a preferred processing service to do data analysis. This model provides users an easy way to set a processing service and get the service outputs.
- **Workflow**: If an analysis task is too complex to be performed by an individual service, this model allows user to build a chain of services to perform the task.
- **Data Visualization**: Users can set up their own preference on how to display data, such as overlay sequence, data subsetting, and image palette. A set of different rendering services is provided for the different data visualization methods.

Generally, a view is a representation of the information in a model. When the relevant model is changed, the view is notified to make corresponding changes by using a "listener" callback method appropriate to the event type. Views may also provide means to send user gestures (e.g. mouse clicks for zoom out) to a controller. Figure 2 is the main view of data visualization. It shows the data structure and image representation. A controller translates interactions with the view into actions to be performed by the model. Controllers modify the information and state of a model. The separation of model and view allows multiple views to be used for the same model. Consequently, it is easy to extend the system because a view and some controller logic can simply be written and wired into the existing application models.

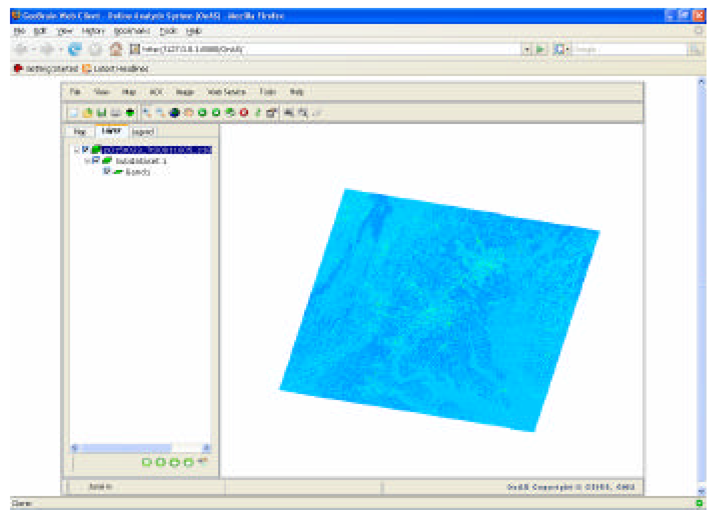

Fig. 2. Main View of Data Visualization

Catalogue Services for the Web (CSW) [2] serves as a directory, helping users to publish meta-information describing geospatial data and services and allowing users to discover the resources needed by querying meta-information.

The Web Processing Service (WPS) provides a domain-specific computational model usable for calculations of all magnitudes, from simple spatial calculation, to global climate change enabling users to do data analysis over the network. The WPSs can also be chained together to perform more complex analysis tasks.

Data Service provides users a common data environment to achieve data interoperability. OGC WCS provides intact multi-dimensional and multi-temporal geospatial data as coverages, to meet the requirements of client-side rendering, input for scientific computational models and clients other than simple viewers. OGC WFS (Web Feature Service) [3] supports the networked interchange of geographical vector data as a feature encoded in GML (Geographic Markup Language). OGC WMS (Web Map Service) [4] provides geospatial data as a map, which is generally rendered dynamically from real geographical data as a spatially referenced pictorial image such as PNG, GIF, or JPEG.

III. CATALOG SERVICES

In a distributed computing environment, a catalog service plays a directory role by using meta-information to help providers advertise the availability of resources and requestors to discover the right resources by querying. There are two leading general models for catalog services: the Electronic Business Registry Information Model (ebRIM) [5] and the Universal Discovery Description and Integration (UDDI) model [6]. For the geospatial domain, ebRIM is more general and extensible because it provides comprehensive facilities compliant with the ISO 11179 set of standards to manage metadata. The OGC CSW, an ebRIM profile for Web-based geospatial catalog services, is an open standard interface dopted by industry consensus, to online catalogs of geographic information and Web-accessible geoprocessing services. CSW specifies interfaces, HTTP protocol bindings, and a framework for defining application profiles required to publish and access digital catalogues of metadata for geographic data, services, and related resource information. Figure 1 shows an extended ebRIM implemented in conformity with ISO 19119:2005 and ISO 19115:2003. The GeOnAS provides the user an OGC CSW interface to register, discover and access a wide variety of distributed geo-resources (i.e. geospatial data, applications, and services) by searching distributed catalog services, such as GEOSS Clearinghouse, NASA ECHO (Earth Observing System Clearinghouse), and GeoBrain catalog.

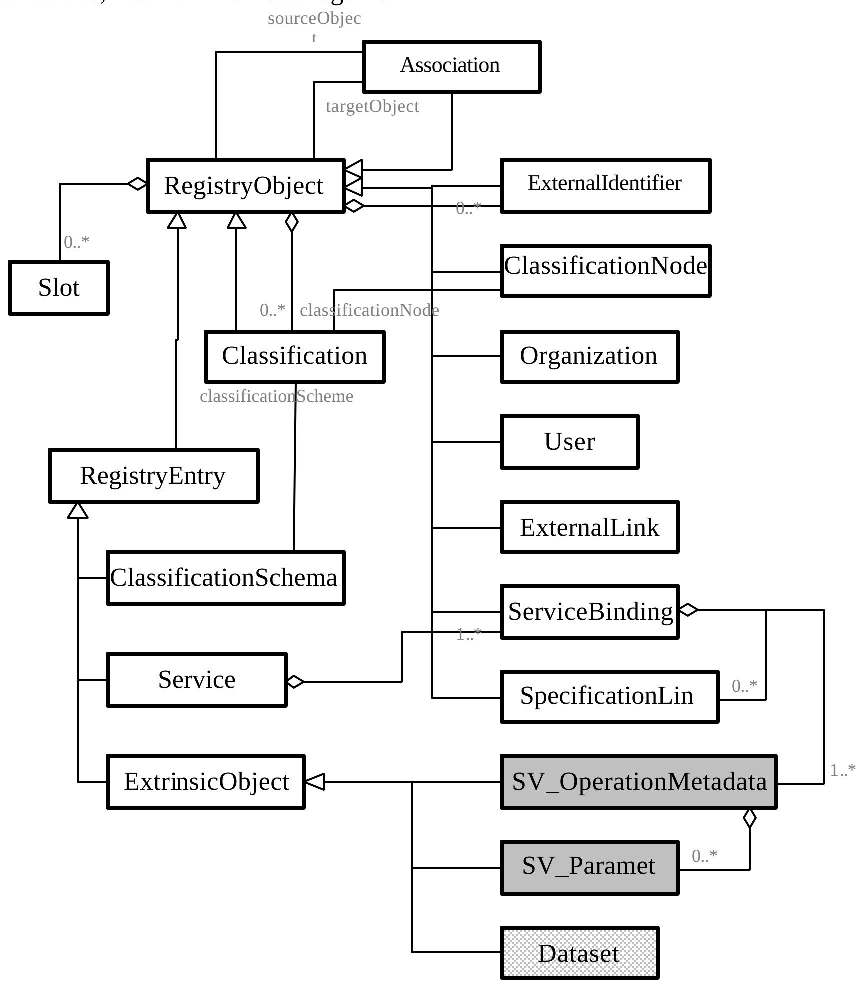

Fig. 3. An Extended ebRIM for Geospatial Data and Services

3.1 Service Registration and Discovery

In ebRIM a class "*Service*", which provides a flexible way to describe a network accessible service offering, is defined. CSW has additional extensions that accommodate the service metadata descriptions in ISO19119:2003, Figure 4 shows the user interface for service registration. The "*Name*", "*Description*", "*Version*" and "*Keywords*" describe the basic attributes of "*Service*". The "*WSDL Address*" specifies the URL of the service's Web Service Description Language (WSDL). WSDL includes all necessary information for "*ServiceBinding*", "*SV_OperationMetadata*", and "*SV_Parameter*", such as operation connect point, binding protocol, and message schema. However, there is no relevant WSDL for most OGC services. In such a case, the "Service Address" is used to specify the service access point directly. Different geospatial services have different application scopes.

Service classification uses standard or proprietary taxonomies ("*ClassificationSchema*") to assign a class ("*ClassificationNode*") to a service, in order to indicate service functionalities explicitly and facilitate service discovery. GeOnAS supports a variety of classification methods to enable the service publisher to identify which of the following domains a service belongs to at publication time:

- **OGC**: the classification schema of the OGC services, such as WCS, WFS, WMS, and WCTS (Web Coordinate Transformation Service).
- **ISO 19119**: the hierarchical classes of services, including human interaction services, model/information management services, workflow services, processing services, communication services and system management services, based on the semantic type of computation defined in ISO 19119:2005,
- **NASA Global Change Master Directory** (GCMD): the hierarchical classes of services defined especially for the sciences, for example environmental advisory services, hazards management services, and model services
- **LAITS_VDP**: a proprietary classification schema of services for virtual data products defined in GeoBrain

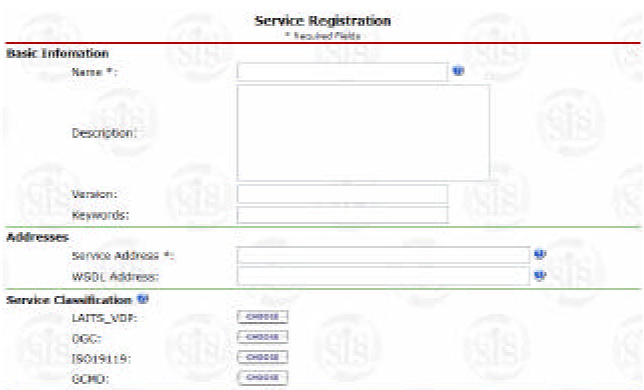

Fig. 4. User Interface for Service Registration

In some cases, service registration also uses "*Association*" to define which operations can operate on a specific dataset or dataset collection.

GeOnAS provides a set of core properties that can be queried to enable cross-catalogue service discovery. They are shown in Table 1. The same queries can be executed against any catalog service without prior negotiation on information content.

Thus, a user can easily discover the desired service instance by specifying a well-known service type, such as OGC WMS. The user can also find an ISO 19119 thematic classification service that can handle MODIS data by combining "Type" and "OperateOn" in a query.

Table 1 Core properties that can be queried for services

Name	Definition	Data Type
ID	Service unique identification	UUID
Name	Service title	CharacterString
Type	Service type	ClassificationSchema
Code	Service taxonomy	ClassificationNode
OperateOn	Coupled data resource	Dataset
All	No constraint	CharacterString

3.2 Data Registration and Discovery

The CSW is further extended to accommodate the metadata prescriptions for geospatial data in ISO 19115:2003. The class "*Dataset*" is added into ebRIM to provide a flexible way to describe a data offering that is network accessible.

In GeOnAS, most data come from OGC data services. Two mechanisms for data registration are implemented in GeOnAS. One is like service registration. It allows the data publisher to use a set of given queryable properties as meta-information to register data. These properties include not only spatial and temporal attributes, but also platform and keyword information defined in ISO 19115:2003, such as instrument name, sensor name, topic keyword, parameter keyword and discipline keyword. The other mechanism is "Harvest". It allows the data publisher to request the catalog service to harvest a data resource from a specified network location by invoking the "getCapabilities" operation and "describe***" operations, thereby realizing a "pull" model for data registration. The "getCapabilities" operation, a common operation of OGC services, responds with service metadata, including the service-specific data information. The WFS "*describeFeatureType*" operation describes the structure of any feature type it can service. The WCS "describeCoverage" operation response gives full descriptions of one or more coverages served by a particular WCS server. By parsing the responses of these operations, all necessary meta-information is registered automatically.

Figure 5 shows the user interface for data discovery based on the following four sets of queryable properties.

- "*Catalog*" (Figure 5-a) allows the user to search distributed catalog services, currently
- "*GMU-LAITS CSW*", which maintains the global Landsat TM/ETM information built by GeoBrain and "*NASA ECHO*", which is a metadata clearinghouse and order broker being built by NASA's Earth Science Data and Information System (ESDIS).
- "*Spatial*" (Figure 5-b) enables the user to provide spatial constraints for data discovery. Google Map is integrated to give the user a convenient way to find a desired location. The user can also select states, counties, and cities within the U. S. or input the country name/code to provide a spatial bounding box for the area of interest.
- "*Temporal*" (Figure 5-c) allows user to query time series data in which "*ProductTime*" specifies the date

and time of data generation and "*CollectionRange*" specifies the date and time the collected data cover.

- "*Others*" (Figure 5-d) allows the user to discover geospatial data through data physical attributes in which "*PlatformShortName*", "*InstrumentShortName*", and "*SensorShortName*" specify, respectively, the short names of platform, instrument, and sensor of the collecting dataset. "*DisciplineKeyword*" and "*TopicKeyword*" describe the general discipline area and topic area of data collection. "*TermKeyword*" indicates the scientific parameters for data collection. "*VariableKeyword*" presents the specific science parameter content. "*ParameterKeyword*" describes high-level specific characteristics of data collection. "*DayNightFlag*" identifies if the dataset is collected during the day, night or both.

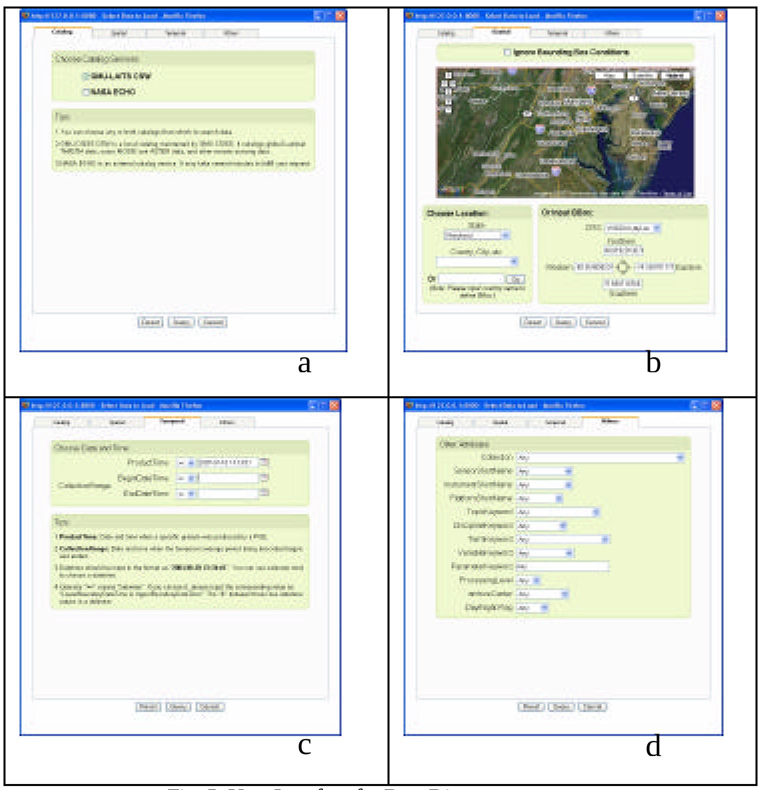

Fig. 5. User Interface for Data Discovery


Fig. 6. Results of Data Discovery

Based on the queryable properties the user selects, the GeOnAS returns a result which contains a data preview, a data description, and information about the data size and format, as shown in Figure 6. Then the user can select the data of interest for online analysis or download it for future use.

IV. WEB PROCESSING SERVICE

According to the definition in [7], the Web Processing Service (WPS) provides client access to pre-programmed calculations and/or computation models that operate on spatially referenced data. The calculation can be as simple as subtracting one set of spatially referenced numbers from another, or as complicated as a global climate change model. For manipulating and analyzing vector and raster geospatial data, GeoBrain provides more than 20 built-in WPSs and more than 50 relevant operations that are developed on top of the Geographic Resources Analysis Support System (GRASS). Users, through GeOnAS, can use these services to analyze data and mine data from any OGC-compliant online data sources. Moreover, the GeOnAS is open and fully extensible. It allows users to integrate their own WPS to do online analysis.

4.1 GRASS Services

Most existing geospatial software was originally developed for desktop applications and not meant for a distributed environment. However, to redesign an entire application from scratch for a Web services environment is clearly an arduous task. A better approach is to modify or add new modules to the existing software to adapt it to a Web services environment so that the full functionalities of the underlying software are exposed in a useful way.

GRASS [8] [9] is an open source GIS software package with over 350 programs and tools for raster and vector data analysis. The basic idea of developing Web services on top of GRASS is to create atomic services based on the GRASS API. However, all commands in the GRASS API are tightly coupled with each other, in contradiction with the Web service design. Some of them have no explicit physical meaning unless they are combined with others in some application. Therefore, a set of services has been developed on top of a large GRASS script which represents a complex physical model. For example, "*create_dummy_location*", "*r_in_gdal*", "*r_mapcalc*", and "*r_out_png*" are used in composing a NDVI service. All these services are self contained and only loosely coupled with each other. Table 2 shows the GRASS services available to implement the basic geospatial manipulations and analysis in GeoBrain.

Table 2. GRASS Services in GeoBrain

Name	Description
Image Color Transformation	This service transforms hue-intensity-saturation (his) color space to red-green-blue (rgb) space, or converts an image from the red-green-blue (rgb) space into the hue-intensity-saturation (his) color space. Each output raster map layer is given a linear gray scale color table.
Image Fusion Brovery	This service performs a Brovey transformation on three multispectral satellite image channels and a panchromatic channel to merge multispectral and high-resolution panchromatic channels.
Image Mosaic	This service mosaics two, three, or four images and extends the color map to the range of all images.
Raster Analysis	This service supports basic raster analysis operations, such as to create buffer zones, recategorize data, fill no-data areas, and expand the edges of non-zero clumps, and thin non-zero cells of the input raster map.
Raster Applications	This service offers applications that use specified functions to implement algorithms, such as to calculate NDVI. New applications are being added.
Raster Calculator	This service provides some simple map arithmetic operations for raster processing, for example, the addition, subtraction, and extraction of raster map layers.
Raster Color Table	This service allows the user to create or modify the color table for a raster map layer.
Raster Data Conversion	This service can be used to convert binary and GDAL-supported raster formats to another format, such as GeoTIFF or PNG.
Raster Image Processing	This service provides necessary operations for image processing, e.g. image blending, compositing or shading.
Raster Map Statistics	This service allows a user to calculate local statistics, global statistics, and univariate statistics to generate profiles and transects, and to create reports.
Raster Query Information	This service reports some basic properties of raster data such as the map's boundaries, resolution, projection, data type, data base location and mapset, the timestamp, history and category number, category labels.
Raster Stream Extraction	This service extracts a stream network from a digital elevation model (DEM) raster using hydrological models in GRASS GIS.
Raster Surface Interpolation	This service provides raster data interpolation. Surface elevation can be generated from a contour map. Two Inverse Distance Weighted (IDW) surface interpolations of raster data are available.
Raster Topographic Parameters	This service involves computation of the DEM and of basic topographic parameters such as aspect, slope, curvatures (profile, tangential, maximum, and minimum) and morphometric features (peaks, ridges, passes, channels, pits, and planes).
Raster Vectorization	This service extracts points, lines or area edge features from raster input data, and converts data to vector format, depending on the type of raster data.
Vector Data Conversion	This service supports vector data format translation, such as conversion from shp to GML.
Vector Feature Extraction	This service extracts vector objects from a vector map with the desired category(ies).
Vector Map Statistics	This service allows the user to calculate local statistics and univariate statistics, test 15 different normalizations for points, generate quadrat counts, sample a raster map at the point locations, and provide length or area sizes.
Vector Network Analysis	This service provides support for vector network analysis. It implements the following algorithms: creation of subnetworks, vector maintenance, iso-distances (from centers), shortest path, optimal route, and optimal connection.
Vector Overlay	This service supports the intersection, union, cutout, overlay or composite of two vector maps
Vector Query Information	This service reports some general properties of vector data such as map

	title, creation date, creator, vector level (topology present or not), number of categories, lines, areas and islands (areas in areas), projection, map boundary coordinates, map scale, further comments, and map attributes (category numbers and labels).
Vector Topology Management	This service creates, changes, or repairs the topology of vector data.
Vector Rasterization	This service transforms a vector map into a raster map.

4.2 Integrating New Services

The most distinguished characteristic of GeOnAS is that it is composed of Web services accessed dynamically to provide the user an open and fully extensible environment. If a user has his/her own geospatial processing services and would like to use them to do data analysis, he/she can integrate these services into GeOnAS to build his/her own GeOnAS. If these services are registered into the catalog, other GeOnAS users will benefit from them. Thereby, GeOnAS becomes more powerful as more users are involved.

To integrate a new service, the service's WSDL is essential. WSDL is an XML-based language that provides a model to describe how to communicate with a Web service [10]. GeOnAS can read the WSDL to determine the operations available in the service, the structures of message types, the binding protocols, and the method of invoking the service. The other requirement for integrating a new service is that the service output must be an accessible network point, i.e. an URL (Uniform Resource Locator). Thus, GeOnAS can parse the URL to import the output data into the system for further use. Integrating a new service to do data analysis consists of the following steps:

1. Register a service with its WSDL, as described in section 3.1.
2. Select data for analysis.
3. Identify a service that can handle the data selected in step 2.
4. Select a proper operation from the list of Figure 7.
5. Input the parameter values of the selected operation, as shown in Figure 8.
6. Invoke the service selected and add the results into the portal, as shown in Figure 9.

Fig. 7. Service Operation List

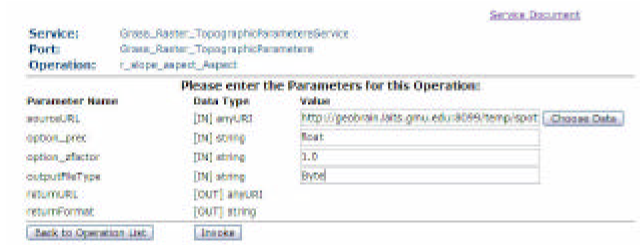

Fig. 8. Parameters of Operation

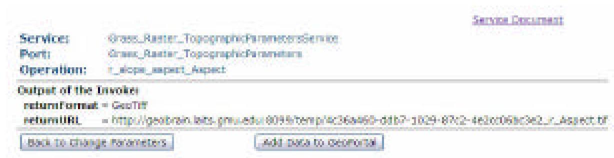

Fig. 9. Output Result of Service Invocation

V. CONCLUSION

The basic functions of GeOnAS are data discovery, data analysis, and data visualization via the Web. GeoBrain itself provides more than 50 geospatial processing functions for manipulating and analyzing vector and raster geospatial data. Through GeOnAS, users can analyze and mine data from any OGC-compliant online data sources with these services. For data visualization, the GeOnAS provides a Web Map Viewer -- a dynamic Web mapping application that spatially locates information to provide visualization of data from disparate data sources and applications. It supports most popular OGC specifications and standards, including OGC context documents, for easy access to previously created maps. In conclusion, GeOnAS provides a new interoperable way of integrating and analyzing distributed heterogeneous ESS data over the web. It is a realization of the Web-and-Service-Centric paradigm in the geospatial domain.

ACKNOWLEDGEMENT

This work is supported by grants from NASA Advanced Information System Technology (AIST) program (NAG5-13409, PI: Dr. Liping Di) and NASA REASoN program (NNG04GE61A, PI: Dr. Liping Di).

REFERENCES

[1] J. Evans, 2003. OpenGIS® Web Coverage Service. OpenGIS Project Document: OGC 03-065r6, Open Geospatial Consortium Inc.
[2] D. Nebert, ed., 2003. OpenGIS® Catalog Services Specification. OpenGIS Project Document: OGC 03-108, Open GIS Consortium Inc.
[3] P. Vretanos, 2002. OpenGIS® Web Feature Service. OpenGIS Project Document: OGC 02-058, Open Geospatial Consortium Inc.
[4] J. de la Beaujardiere, 2004. OpenGIS® Web Map Service. OpenGIS Project Document: OGC 04-024, Open Geospatial Consortium Inc.
[5] OASIS, 2002. "OASIS/ebXML registry information model v2.0". http://www.oasisopen.org/committees/regrep/2.0/specs/ebRIM.pdf.
[6] OASIS, 2000. "The UDDI technical white paper". http://uddi.org/pubs/uddi-tech-wp.pdf
[7] P. Schut and A. Whiteside, 2005. OpenGIS® Web Processing Service. OpenGIS Project Document: OGC 05-007r4, Open GIS Consortium.
[8] GRASS Website. http://grass.itc.it
[9] H. Mitasova, M. Neteler, 2002. Open Source GIS: A GRASS GIS Approach, 460 pages, Kluwer Academic Pres
[10] W3C, 2001. Web Services Description Language (WSDL) 1.1. http://www.w3.org/TR/wsdl