DEVELOPING A CLOUD-BASED ONLINE GEOSPATIAL INFORMATION SHARING AND GEOPROCESSING PLATFORM TO FACILITATE COLLABORATIVE EDUCATION AND RESEARCH

Z. L. Yang [a, *], J. Cao [a], K. Hu [a], Z. P. Gui [b], H. Y. Wu [a], L. You [c]

[a] The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China – (yangzelong, caojun1212, hukai, wuhuayi)@whu.edu.cn
[b] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, China – zhipeng.gui@whu.edu.cn
[c] Faculty of Computer Science and Information Engineering, Hubei University, Wuhan, China – yoyo@huhu.edu.cn

KEY WORDS: GeoSquare, geospatial information resources, collaborative, education, research, Online GeoProcessing, Cloud Computing

ABSTRACT:

Efficient online discovering and applying geospatial information resources (GIRs) is critical in Earth Science domain as while for cross-disciplinary applications. However, to achieve it is challenging due to the heterogeneity, complexity and privacy of online GIRs. In this article, GeoSquare, a collaborative online geospatial information sharing and geoprocessing platform, was developed to tackle this problem. Specifically, (1) GIRs registration and multi-view query functions allow users to publish and discover GIRs more effectively. (2) Online geoprocessing and real-time execution status checking help users process data and conduct analysis without pre-installation of cumbersome professional tools on their own machines. (3) A service chain orchestration function enables domain experts to contribute and share their domain knowledge with community members through workflow modeling. (4) User inventory management allows registered users to collect and manage their own GIRs, monitor their execution status, and track their own geoprocessing histories. Besides, to enhance the flexibility and capacity of GeoSquare, distributed storage and cloud computing technologies are employed. To support interactive teaching and training, GeoSquare adopts the rich internet application (RIA) technology to create user-friendly graphical user interface (GUI). Results show that GeoSquare can integrate and foster collaboration between dispersed GIRs, computing resources and people. Subsequently, educators and researchers can share and exchange resources in an efficient and harmonious way.

1. INTRODUCTION

With the advancement of sensors and information technologies, numerous GIRs, including geodata, algorithms, applications and models, are now available online for public use. Experts from other disciplinary can use it to conduct experiments. However, traditional geographical information system is cumbersome and expensive. Web geoprocessing based on service collaboration has been receiving increased attention for geoscientific knowledge discovery (Kiehle, Greve, and Heier, 2007). And collaborative geoprocessing models are becoming one of the major solutions to significantly enhance the capacity to derive geo-information and knowledge over a network (Zhao et al, 2012). At the same time, the emergence of the cloud computing is promoting a transformation from traditional desktop geoprocessing to distributed collaborative geoprocessing (Wu et al, 2010). Therefore, this article proposes to develop a cloud-based online geospatial information sharing and geoprocessing platform to facilitate collaborative education and research. However, there are still some challenges: the heterogeneity and complexity hinders the discovery and sharing of GIRs; the cross-domain barriers restrict users without specialized knowledge; the privacy of GIRs puts forward higher requirement on the development of platform. To address these problems, this article utilizes a set of methodologies: (1) GIRs registration and multi-view query functions allow users to publish and discover GIRs more effectively. (2) Online geoprocessing and real-time execution status checking help users process data and conduct analysis without pre-installation of cumbersome professional tools on their own machines. (3) A model sharing mechanism is

proposed to share domain experts' domain knowledge with community members. (4) Platform provides three levels of resource management and sharing scopes, including public, group and private, to ensure the privacy of users' GIRs and other information.

The remainder of this article is organized as follows. Section 2 discusses related work. Section 3 introduces the architecture of the GeoSquare. Section 4 introduces the key technologies that enable the GIRs to work collaboratively. Section 5 describes the implementation and typical applications. Section 6 concludes with results and discusses future work.

2. RELATED WORKS

Geospatial information sharing and geoprocessing, which can promote the on-demand instant transformation of geodata into knowledge in the web environment, have attracted attention. As a result, there are many GIRs published online as web services which can be used for collaborative education and research in cross-domain. Gui et al. (2013a) designed and developed GeoSearch which adopts brokering middleware to integrate heterogeneous resources, batch processing, asynchronous messaging and multi-threaded-based search mode to improve performance, a semantic-based query to improve recall level and precision, sophisticated functionalities to improve user experience and assist decision-making. Gui et al. (2013b) used Rich Internet Application(RIA) technologies to assist geospatial resource discovery through the GeoSearch. You et al. (2012) proposed a framework for geospatial web service composition that supports real-time status monitoring, which can help users

find the bottleneck of efficiency of services chain. Gui et al. (2008) proposed a novel data-dependency directed graph and block structures based abstract geospatial information service chain model. Wu et al. (2011) developed GeoChaining that integrates visual editing, validation, deployment, execution etc. into a unified platform and supports reuse of model. Wu et al. (2014) presented a fully asynchronous and status-tracking pattern geoprocessing service orchestration (Fast) for collaborative geoprocessing. Wu et al. (2015) proposed a cloud-based framework for building, executing and sharing collaborative geoprocessing models and migrated the system into cloud. Gong et al. (2012) proposed the concept of geospatial services web (GSW) to underpin the development of a future geospatial cyberinfrastructure. Qi et al. (2015) extended the WSDL to include specifications and constraints and dynamically generate GUI for workflow invocation and results' representation. Qiu et al. (2012) developed a GRASS Web Application Software System (GWASS) that allow users to call functions in the platform as desktop software. This work provides a cloud-based online geospatial information sharing and geoprocessing platform to facilitate collaborative education and research.

3. ARCHITECTURE

Aiming to facilitate online geospatial information sharing and geoprocessing, GeoSquare was implemented based on a cloud-based framework. The architecture (Figure 1) consists of three tiers: the application tier, computation tier and resource tier. In application tier, GIRs registry center and user profiles were designed for GIRs registration and management, while five geoprocessing components including model builder, model executer, model monitor, result visualizer and model publisher are in charge of collaboratively online geoprocessing. The computation tier involves collaboration engine pool and status collector. The engine pool is composed of several engine servers where the geoprocessing models are deployed and executed through WS-BPEL workflows. Status collector captures runtime information and update continually when executing models. All GIRs and their metadata are managed and organized in resource tier, which constitutes the basis of the framework. Newly created and already existing models are stored in the model repository and shared among engine servers.

The whole framework is built on Microsoft Azure Cloud, which provides abundant cloud storage and computing resources for hosting applications and services. To enhance the flexibility and capacity of GeoSquare, some features in Azure Cloud were utilized, i.e., Azure load balancer (ALB) and Azure blob storage. ALB is a load-balancing mechanism that acts as a proxy and automatically distributes network or application traffic across multiple Azure-hosted servers. When executing a geoprocessing model, the ALB will distribute the invocation to an available engine server which is idle. The engine pool can scale up as the concurrency increases. Azure blob storage provides elastic storage and acts as a file system with unlimited capacity. All GIRs from user workplaces were stored in Azure blob storage.

4. METHODOLOGIES

4.1 GIRs registration and multi-view query

GIRs can be registered in registry center and queried by metadata. They are divided into three groups including geodata (e.g., vector data and image data), services (i.e., W3C web services and OGC WPS services) and models (i.e., the unit of a series of geoprocessing services). The registry center provides multi-view approaches including snapshot, list and chart for users to query GIRs by various filters (e.g., geoextent, data size, data type). The registration and multi-view query allow users to publish as well as discover GIRs more effectively.

4.2 Online GeoProcessing and real-time execution status checking

Geoprocessing can be executed online either using atomic services or geoprocessing models and the execution status information can be shown dynamically for users. Pre-installation of cumbersome professional tools on user machines are not required since the process is implemented on server-side. When executing a geoprocessing model, model executer is used to remotely invoke the geoprocessing model deployed in the engine servers. Moreover, the model monitor captures the runtime information of the geoprocessing model periodically from the status collector and displays them to users. Once the geoprocessing is finished, users can conveniently browse the process results with result visualizer according to the result types. For example, Google Earth is integrated for previewing the image data and charts could be provided for statistical results.
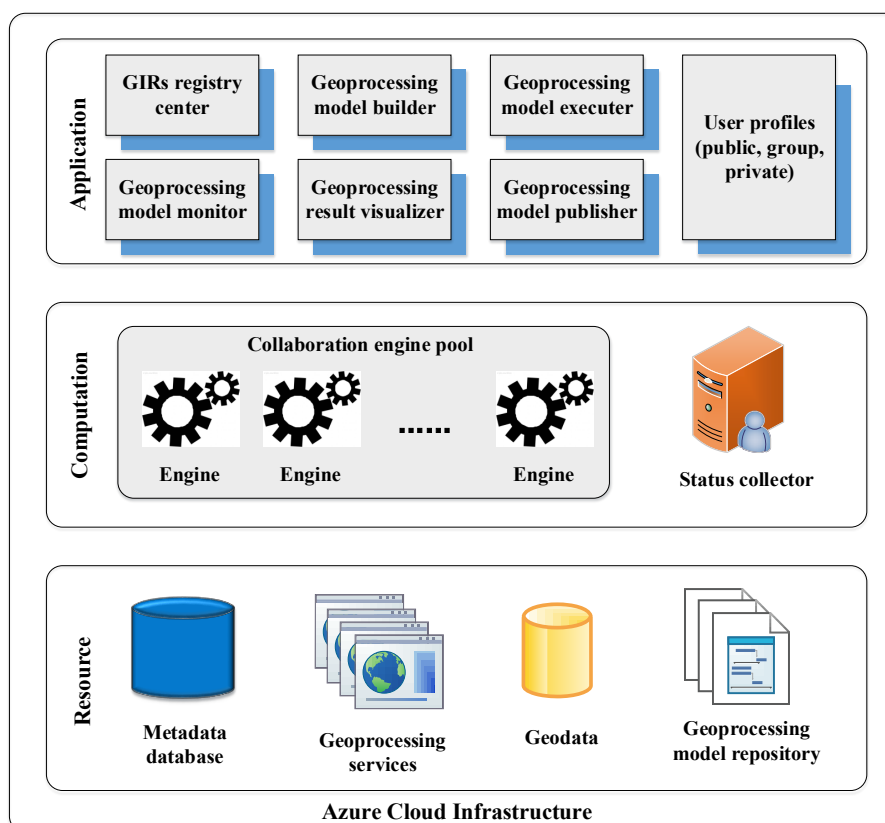
Figure 1. Architecture of GeoSquare

4.3 Service chain orchestration

For domain experts, they can contribute and share their knowledge with community members through model building and publishing by using model builder and model publisher respectively. The model builder provides a graphic user interface (GUI) to build new processing models or edit old ones. Users specify the structure and properties of a model. When a model is created, it can be stored in the model repository for reuse. Users can either create a new model or use an old one. Besides, the newly created geoprocessing models must be published on the engine server for invocation through model publisher, the model metadata is registered in the registry center.

4.4 User inventory management

GeoSquare allows users to collect and manage their own inventory conveniently. Users can upload their geodata and download the geoprocessing historical data. The registry center provides three levels of resource management and sharing scopes including public, group and private. Users can deliver the geodata, services and models to the public if they want. A group can also be defined by users to share GIRs within a specific scope. When users want to keep their inventory only available to their own, the private definitions could be executed. The user inventory can be managed flexibly according to different scopes that users require.

5. IMPLEMENTATIONS AND APPLICATIONS

GeoSquare was developed on the Microsoft Azure Cloud infrastructure. The application tier was developed on Eclipse 6.0 IDE. The registry center was built with Ext Google Web Toolkit

(GXT). The geoprocessing model builder was developed with Java RCP techniques. The geoprocessing result visualizer was developed using Google API. The ActiveBPEL engine was taken as the collaborative engine for the geoprocessing models. All GIRs were stored in the Azure blob storage while their metadata was managed in SQL Azure.

To support interactive teaching and training, the Rich Internet Application (RIA) technology was adopted to create user-friendly Graphical User Interface (GUI). The main GUI is shown as Figure 2. The left panel is the registry center where GIRs were presented as a tree catalogue. The right panel includes map viewer and metadata viewer that displays detail metadata and a snapshot of resources in registry center. Several typical use-cases were developed for Remote Sensing (RS) and Geographic Information Systems (GIS) teaching and training, such as, image classification (Figure 3(a) and Figure 3(b)) and quadrat analysis (Figure 3(c) and Figure 3(d)).

6. CONCLUSIONS AND FUTURE WORKS

This article presents the framework and methodologies of GeoSquare, a platform aiming to facilitate collaborative education and research by utilizing online geospatial information sharing and geoprocessing.

(1) To satisfy users' multiplex search demand and improve user interactivity, registration and multi-view query functions are achieved. Users can publish and discover GIRs more effectively.

(2) To broad the range of platform users, online geoprocessing and real-time execution status checking avoid pre-installation of cumbersome professional tools on users' machines. Common

users can directly use the geodata and geoprocessing to conduct experiments through web browsers.

(3) To simplify operations and share experts' knowledge, a model sharing mechanism is proposed, which enables domain experts to contribute and share their domain knowledge with community members through workflow modelling.

(4) To protect the privacy of registered users, a user inventory management mechanism is utilized, which provides three levels of resource management and sharing scopes, including public,

group and private. It allows registered users to collect and manage and their own GIRs, monitor their execution status, and track their own geoprocessing histories.

For additional improvement in flexibility and enriching the functionality of platform, future research would include: (1) Provide and consume web service standard and casting interfaces to facilitate interoperation with other systems. (2) Clarify the GIRs classification.
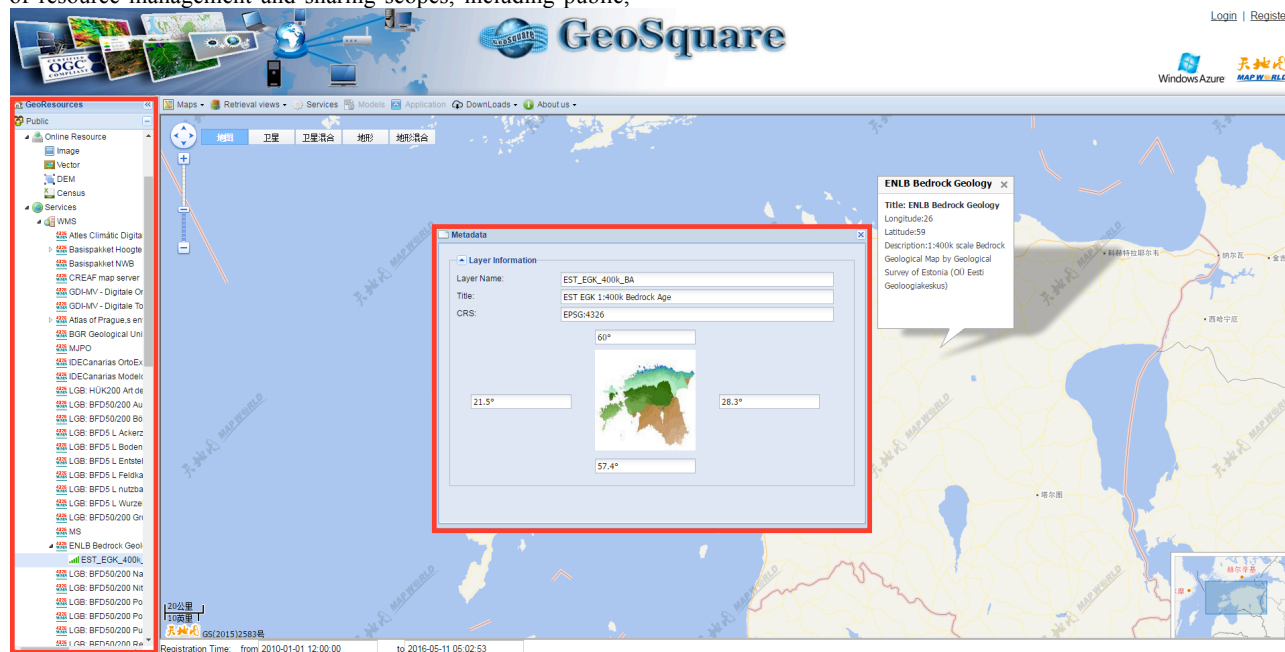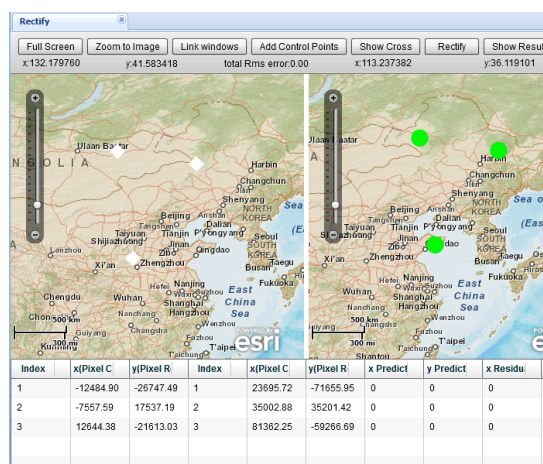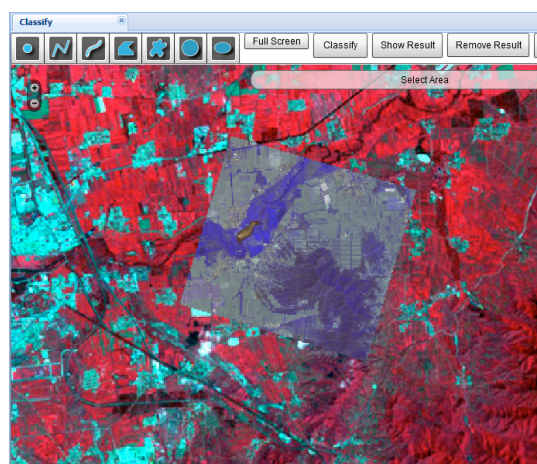

Figure 2. GUI of GeoSquare

(a) Image rectification
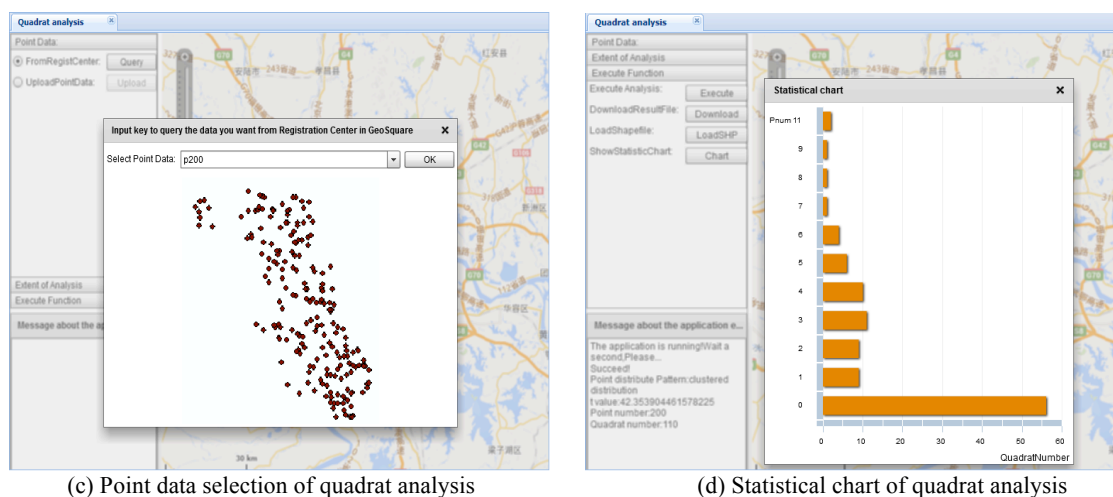

(b) Image classification

(c) Point data selection of quadrat analysis   (d) Statistical chart of quadrat analysis

Figure 3. RIA-based Environment for RS and GIS Teaching and Training

REFERENCES

Endpoints, A., 2004. ActiveBPEL, http://www.eclipse.org/bpel/.

Gong, J., Wu, H., Zhang, T., Gui, Z., Li, Z., You L., 2012. Geospatial Service Web: towards integrated cyberinfrastructure for GIScience. *Geo-spatial Information Science*, 15(2), 73-84.

Google, 2012. GXT, https://www.sencha.com/products/gxt/.

Gui, Z., Wu, H., Wang, Z., 2008. A Data Dependency Relationship Directed Graph and Block Structures Based Abstract Geospatial Information Service Chain Model, In: *Proceedings of the 2008 Fourth International Conference on Networked Computing and Advanced Information Management*, 2, 21-27.

Gui, Z., Yang, C., Xia, J., Liu, K., Xu, C., & Jing, L. P. L. 2013a. A performance, semantic and service quality-enhanced distributed search engine for improving geospatial resource discovery. *International Journal of Geographical Information Science, 27*(6), 1109-1132.

Gui, Z., Yang, C.*, Xia, J., Li, J., Abdelmounaam, R., 2013b. A visualization-enhanced graphical user interface for geospatial resource discovery. *Annals of GIS*, 19(2), 109-121.

Kiehle, C., Greve, K., & Heier, C. 2007. Requirements for next generation spatial data infrastructures - standardized web based geoprocessing and web service orchestration. *Transactions in GIS, 11*(6), 819-834.

Qi, K., Gui, Z.*, Li, Z., Guo, W., Wu, H., Gong, J., 2015. An Extended Mechanism to Verify, Constrain and Enhance Geoprocessing Workflows Invocation. *Transactions in GIS*.

Qiu, F., Ni, F., Chastain, B., Huang, H., Zhao, P., & Han, W., et al. 2012. Gwass: grass web application software system based on the geobrain web service. *Computers & Geosciences, 47*(4), 143-150.

Wu, B., Wu, X., & Huang, J. 2010. Geospatial data services within Cloud computing environment. *Audio Language and Image Processing (ICALIP), 2010 International Conference on* (pp.1577-1584). IEEE.

Wu, H., You, L., Gui, Z., 2011. DIY GEOSPATIAL WEB SERVICE CHAINS: GEOCHAINING MAKE IT EASY. In:

ISPRS joint-WG workshop on 'Geospatial Data Infrastructure: from data acquisition and updating to smarter services', Guilin, China, 2011, Vol. XXXVIII-4/W25.

Wu, H., You, L., Gui, Z., Gao, S., Li, Z., Yu, J., 2014. FAST: A Fully Asynchronous and Status-Tracking Pattern for Geoprecessing Services Orchestration. *Computers & Geosciences*, 70, 213-228.

Wu, H., You, L., Gui, Z., Hu, K., & Shen, P. 2015. Geosquare: collaborative geoprocessing models' building, execution and sharing on Azure Cloud. *Annals of GIS*, 21(4), 287-300.

You, L., Gui, Z., Guo, W., Shen, S., Wu, H., 2012. A geospatial web services composition framework supporting real-time status monitoring. *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, I-4, 175-179.

Zhao, P., L. Di, and G.Yu. 2012. Building Asynchronous Geospatial Processing Workflows with Web Services. *Computers & Geosciences*, 39: 34–41.