A Centralized System for Managing, Archiving and Serving Scientific Data in Ecohydrological Research

Mutao Huang

College of Hydropower & Information Engineering, Huazhong University of Science and Technology, Wuhan, 430074, China
hmt1973@sina.com

Yong Tian[*]

School of Environment Science and Engineering, South University of Science and Technology of China, Shenzhen, 518055, China
95297626@qq.com

Abstract—**Transboundary and cross-catchment access to ecohydrological data is the key to designing successful environmental policies and activities. Access to and visual representation of scientific information for the public is essential for attracting greater awareness of ecosystem matters. This requires a centralized system to bring together diverse data so that they can be integrated and analyzed in a single environment. This paper presents the efforts of combining web services and ontologies to address these challenges. In this study, a centralized system for managing, archiving and serving scientific data in ecohydrological research was designed and developed. The system enables to mash-up both observations and modeling results together. The applicability of the system is demonstrated through a case study in the Heihe River Basin located in northwestern China.**

Keywords—Web services; ontologies; Service-oriented architectures; ecohydrology

I. INTRODUCTION

In the past 10 years, using the watershed as the unit to establish an integrated watershed observing system has been an important trend in integrated eco-hydrologic studies. A number of terrestrial research programs have used watersheds as the basic unit for research and observation, including the Terrestrial Environmental Observations[1,2] (TERENO), the Critical Zone Observatory[3] (CZO) proposed by the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI), and the Danish Hydrological Observatory (HOBE) [4].

In China, a similar observation program was implemented in the Heihe River Basin (HRB) from 2010. The HRB has served as a testbed for the ecohydrology study since the 1980s. A large number of research programs have been implemented in the HRB from a variety of disciplines such as hydrology, physical geography, ecology, land surface study, economic geography, and remote sensing, producing rich knowledge and data about the basin. In particular, a comprehensive ecohydrological experiment called HiWater (Heihe Watershed Allied Telemetry Experimental Research) is being implemented[5]. A large amount of high-quality and high-resolution data have been produced by this experiment. The data sets cover near all aspects of hydrologic sciences. Many efforts have been devoted to distribute these data to the researchers and the public. However, many of these ecohydrological data products have not been used to their full potential, and are not used in conjunction with one another. Furthermore, there has been no successful system to integrate this disparate data at a basin scale.

This research is motivated by the need to integrate and explore the various sources of ecohydrological data that have been made publicly available through web services. In particular, we focus on a use case where a hydrologist would like to estimate water balance and ecohydrological processes for a watershed to assess the ecological response of a watershed to hydrology conditions. The results of this type of analysis can be used to explore relationships between the hydrology and ecology processes. Currently, the users have to download these data from different sources, integrate them together and then analyze the data using specialized software.

The work presented in this paper offer users interactive dynamic web applications that enable retrieving both observational and modeling data in a single environment. This paper presents the design and development of a Service Oriented Architectures (SOA) and Rich Internet Applications (RIAs) based system meets the challenges in discovering and accessing scientific data in the realm of the ecohydrology domain.

II. THE ECOONLINE SYSTEM

A. System overview

The EcoOnline is a distributed, service-oriented architecture system. The design for the system is shown in Fig. 1. Components and services in each tier perform similar tasks with the accordance to the purpose of the tier in which they reside. The system is designed based on SOA, which is a conceptual architecture using loosely coupled and interoperable web services in a distributed computing environment.

The bottom tier is responsible for data storage and computation. The system utilizes the most widely used OGC protocols including WMS, WFS and WCS to discover and deliver image map, vector map, and coverage information. Meanwhile, the discovery and retrieval of observational data is accomplished by leveraging native web services.

The middle layer is the service layer which is the most important part of the whole architecture. The design and implementation of this layer is based on our previous work. The layer consists of several facilities, including the Enhanced

Service Catalogue (ESC), the domain ontology repository and the service registry. These facilities form the backbone of the entire system. The functionalities of the facilities are exposed through a set of web services. The ESC is an UDDI-based registry that holds metadata about services including technical and semantic information about the services. It provides the interface between the Internet and the web server. Core functionalities of the registry are exposed by the Query Services, which provide access to inquiring the registered services. To alleviate the semantic heterogeneity issues between data repositories, ontology is employed to provide explicit and machine-understandable conceptualizations of water information and semantically annotate the information content of web services. By annotating the services with the semantics contained in the ontology and invoking UDDI APIs, the services matchmaker has the ability to perform semantic based service discovery. The use of a web service API allows the web client to be developed in a modular fashion. It also allows developers to access the system to create their own mash-ups and clients.

The top layer separates the web application interface layer from the service layer, exposing as much of the functionality as possible through an application programmers interface to enable direct client access and interoperability with other systems. With the user-friendly public browser accessible graphical user interface, the web portal provides the users with an integrated view for data interaction, real-time access to online data, data mashup, data download, as well as visualization. A suite of client modules have been developed for the system using RIAs technologies.
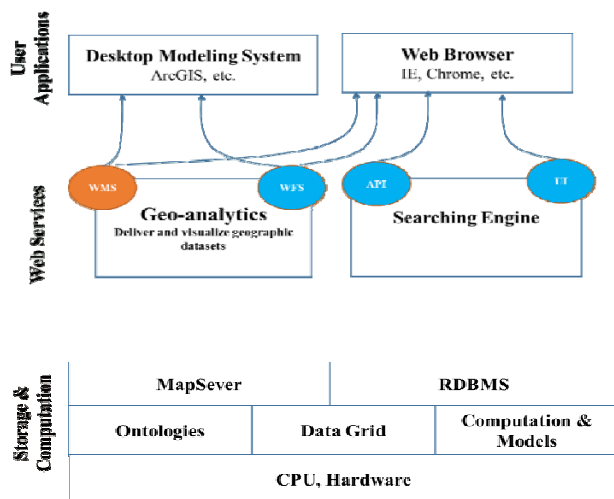

Figure 1. System architecture

B. Domain Ontologies

Ontologies are especially useful because they support the exchange and sharing of information. Domain ontologies are the cornerstone of informatics systems[6,7]. In this work, the ontology mainly serves as a classification system (or a taxonomy) and assists in performing semantic expansion. In this work, the ontology mainly serves as a classification system

(or a taxonomy) and assists in performing semantic expansion. Our semi-automatic ontology construction process for ecohydrological domain comprises several phases, including collection and analysis of domain information, documents pre-processing, concept extraction, relation extraction, ontology formalization, ontology evaluation, etc. Fig. 2 shows the user interface used to perform the procedure. Fig. 3 shows an example of domain ontology constructed by the tool.
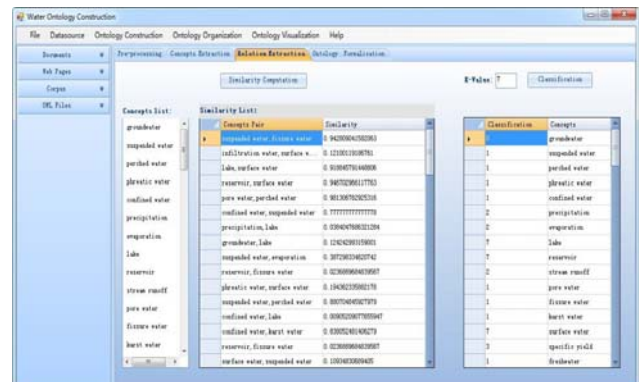

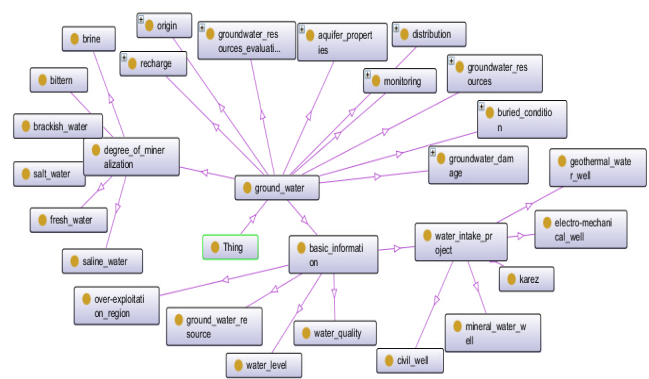Figure 2. Semi-automatic tool used to construct ecohydrological ontologies

Figure 3. Example of domain ontology related to groundwater sub-domain.

C. Data sets Extraction and Calculation

Data Extraction and Calculation (DEC) is the process of extracting data from external sources, transforming the data based on system requirements, and loading the data into a data warehouse. The processes in the DEC workflow consist of the steps needed to extract the watershed boundaries, data grids, mask grids, and basin-averaged time series. A conceptual overview of the processing steps needed are shown in Fig. 4.
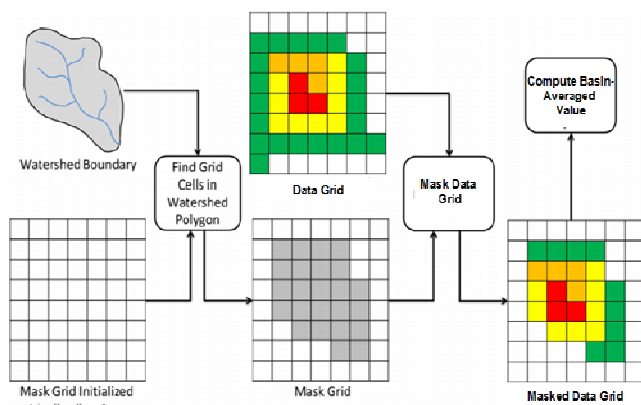
Figure 4. Conceptual overview of spatial processing steps to create basin averaged variable time series

III. CASE STUDY

A. Study area

The study area selected in this study is the Heihe River Basin (HRB). The HRB is the second largest inland river basin in China. As located in a transition zone between a semi-arid zone (the Qinghai-Tibet Plateau) and an arid zone (the Mongolian Plateau), the HRB is one of most water-stressed basins in China. Rapid economic development has come at the expense of degradation of ecosystem in this area. Water shortages and irrational water exploitation have caused a number of ecological problems and social conflicts, including reduction of flow discharged to downstream, gradual disappearance of wetlands and some seasonal rivers, desertification, salinization, and shrinkage of the terminal lakes (i.e., East and West Juyan Lake). In short, typical ecological problems that often appear in semi-arid areas around the world are all found in the HRB.

A world-class ecohydrological network has been established in the HRB. A set of super high-resolution airborne remote-sensing data has also been obtained. Furthermore, many modeling products have been generated by many researching groups. These data include remote sensing products (e.g., ET, LAI, etc.), point observation time series (e.g., hydrography, groundwater level, water temperature, etc.), integrated models (i.e., model grids and parameters) and 3D/4D simulation results generated by the integrated models. Improved estimation approaches for the key hydrological variables, i.e., precipitation, ET, SM, SWE and groundwater storage. Time series of basin scale remote sensing and assimilated data products, e.g. ET, SM and SWE over both study areas from 2012 to 2015.

B. User interface

The user interface of the EcoOnline was developed by leveraging Silverlight technique. Silverlight is Microsoft's solution for delivering rich, cross-platform interactive experiences for the Web and beyond. The advantages of Silverlight are numerous. In particular, Silverlight has the advantage of integrating well with other Microsoft .NET based

technologies, especially with Windows Communication Foundation (WCF). WCF is an API in .NET framework for building service-oriented applications. It enables developers to quickly build distributed applications under SOA implementation, and can be easily consumed by Silverlight clients. In this work, the native services aforementioned were deployed by means of WCF RIA services. Another powerful API called ArcGIS API for Microsoft Silverlight was utilized for the integration and visualization of geospatial data. The API also provides some key benefits, such as smoothly and rapidly rendering large amounts of geospatial data and providing native support for performing common GIS tasks.

Fig. 5 shows the main graphic user interface (GUI) of the EcoOnline system. The GUI consists of a map viewer, a searching box and menus. The map viewer plays a core role in displaying GIS data, in which the study area is shown. The user can select one of these layers as base map (e.g., World Imaginary, Physical World, etc.) to attain a unified spatial background context. In addition, the user can freely switch the base map as desired based on personal taste and quality of maps in their local area.

Figure 5. User interface of the EcoOnline system

C. Determination of model structure

There are several potential use cases for the EcoOnline system that take advantage of its unique capabilities. Data providers can use the system's ability to integrate and plot data from multiple sources. For example, users could plot streamflow from a local gauging station. Alternately, local organizations that maintain a network of environmental sensors, such as a state could adapt the centralized system to extract, transform, and load data from their network of sensors into a centralized database. Fig. 6 shows an example of publishing streamflow measured at all the gaging stations in the HRB. The time series viewer floating on the map viewer shows the daily streamflow measured at the Gaoya station.
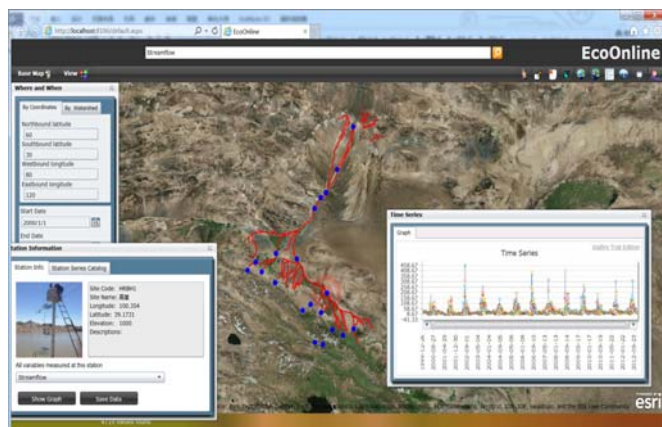
Figure 6. EcoOnline interface showing all the streamflow gaging stations and a hydrography measured at the Gaoya station

Another major use case for the EcoOnline system is a scenario where the modeler would like to publish their gridded modeling results onto this centralized system. Different modelers can publish their results into the centralized database and compare their results. Fig. 7 shows such an example. A gridded evapotranspiration products that are produced by a Remote Sensing based model is published on the EcoOnline system and displayed in the map viewer. When publishing observations or modeling results, different users may have their own terms, but by using ontologies as shared terms, semantic heterogeneities could be avoided.
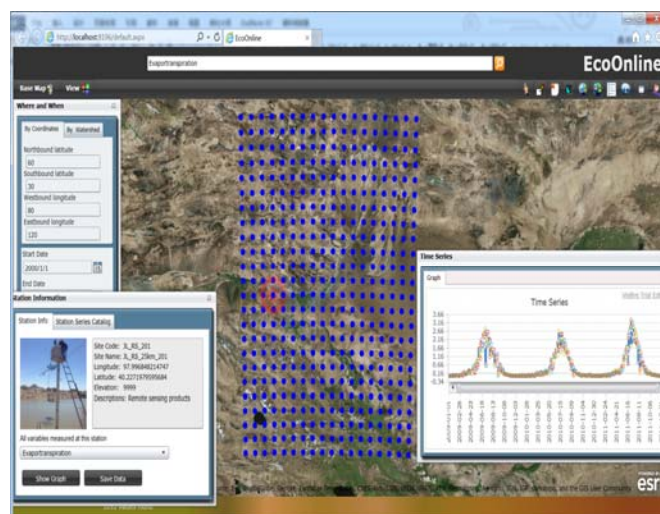

Figure 7. Daily evapotranspiration derived from a Remote Sensing based model

REFERENCES

[1] Baroni, G., S.E. Oswald. A scaling approach for the assessment of biomass changes and rainfall interception using cosmic-ray neutron sensing. J. Hydrol., 2015, 525:264-276.

[2] Qu, W., H.R. Bogena., J.A. Huisman, J. Vanderborght, M. Schuh, E. Priesack and H. Vereecken. Predicting sub-grid variability of soil water content from basic soil information. Geophysical Research Letters, 2015, 2014GL062496.

[3] Melissa A. F., Robert S.A., et al. Hillslope lowering rates and mobile-regolith residence times from in situ and meteoric Be-10 analysis, Boulder Creek Critical Zone Observatory, Colorado. Geological society of America bulletin, 2015, 127:9-10.

[4] Guzinski R., Anderson M.C., et al. Using a thermal-based two source energy balance model with time-differencing to estimate surface energy fluxes with day-night MODIS observations. Hydrology and earth system sciences, 2013, 17:2809-2825.

[5] Li X., Cheng G., et al. Heihe Watershed Allied Telemetry Experimental Research (HiWATER): Scientific Objectives and Experimental Design. Bull. Amer. Meteor. Soc., 94:1145-1160.

[6] ZHENG Ye-lu , HE Qi-yun, QIAN Ping and LI Ze , "Construction of the Ontology-Based Agricultural Knowledge Management System," vol. 11, May 2012, pp. 700-709 .

[7] CarlaFariaa, IvoSerrab, RosarioGirardib. "A domain-independent process for automatic ontology population from text," vol . 95, December . 2014, pp. 26-43, doi:10.1016/j.scico.2013.12.005

.