AWRA 2012 SPRING SPECIALTY CONFERENCE
New Orleans, Louisiana
March 26 – 28, 2012

DEVELOPMENT OF A LIGHTWEIGHT HYDROSERVER AND HYDROLOGIC DATA HOSTING WEBSITE

Jiří Kadlec and Daniel P. Ames*

ABSTRACT: The CUAHSI HydroServer data publishing platform (hydroserver.codeplex.com) is being used by an increasing number of universities and research groups for sharing hydrologic time series data through standards-compliant WaterML web services. Setting up a HydroServer typically requires extensive computer infrastructure including a fully operational Microsoft Windows-based Internet Information Services (IIS) server. This presentation introduces a simplified and streamlined HydroServer implementation that can be deployed using minimal server resources (i.e. on a commercial web hosting platform with restricted server-side access). This "HydroServer Lite" includes all of the requisite components for uploading data in a number of data formats, exposing data for download using WaterML, and registering data services on the HIS Central Catalog. This presentation will include discussion of the design and implementation of the HydroServer Lite system, as well as implementation of the system on a freely accessible web site. The new web site is being developed as an easy-to-use web hosting service for sharing hydrologic data without requiring users to install their own server. After setting up an account on the web site, users can upload and edit data using a web browser interface. The observation data values can be entered in the web browser or uploaded from a variety of text file formats. A web map interface can be used to locate the observation sites in the map. After uploading a time series to the web, the time series is automatically registered at HIS Central catalog and becomes available for searching, downloading and detailed analysis in the free HydroDesktop client software application.
KEY TERMS: hydrologic information system; web services; WaterML; Web 2.0

INTRODUCTION

The Consortium of Universities for Advancement of Hydrologic Science (CUAHSI) hydrologic information system (HIS) is an open, distributed, web based system for sharing hydrologic, climate and water quality time series observation data (CUAHSI, 2011). The system consists of a distributed network of servers, each of which exposes a standard WaterOneFlow web service to access the time series observation data and metadata. This web service is accessible to any user or machine connected to the internet. The web service from any server can be registered in a centralized catalogue (HydroCatalog) and accessed by a number of desktop based and web based client viewer tools. Examples of these clients include HydroExcel, HydroDesktop (Ames et al., 2011) or WRDDs (Huang et al, 2011). Registering a server at the catalogue is not mandatory for making the data accessible to the client. If users know the URL of the web service, then they can enter this URL in the client and use the client to retrieve, access and analyze the time series data from the server.

*Respectively, Graduate Student, Associate Professor, Geospatial Software Lab, Idaho State University, 995 University Blvd, Idaho Falls, ID, 83401, Phone: 208-313-8076, email: kadljiri@isu.edu

The WaterOneFlow web service specification is a platform independent standard (Maidment et al., 2011). The only requirement for a standards-compliant WaterOneFlow web service is to implement the required web methods (GetSites, GetSiteInfo, GetVariables, GetVariableInfo, GetValues) and to ensure that the outputs of these methods are in a valid WaterML format. WaterML is a standard extensible markup language (XML) schema for publishing time series observations. Several technical implementations of the WaterOneFlow web service have been developed. Among the most widely adopted solutions are HydroServer and PyHIS.

The CUAHSI HydroServer (Horsburg et al, 2009) is a comprehensive software tool for management and publishing of hydrologic observations including hydrologic simulation results and water quality data. It requires a virtual private server or dedicated server with the Windows server operating system and the Microsoft .NET framework. The advanced security setting options of HydroServer require it to run as a full trust application. Internally, HydroServer uses the Microsoft SQL server database management system and the Observations Data Model (ODM) database schema (Horsburg et al, 2008). A set of data abstraction layer classes translates the relational schema into the WaterML XML schema objects. The source code is accessible free of charge on the hydroserver.codeplex.com community website. Shipped with HydroServer comes a suite of optional components for data management and visualization: HydroServer Website, Map and web map service (requires an ArcGIS server license), ODM data management tools, ODM data loader and time series analyst. The advantage of HydroServer is high scalability, detailed documentation and rich functionality.

WoFPy (WaterOneFlow - Python) is a platform-independent and database-independent implementation of the WaterOneFlow web service. It is written in the Python programming language. With WoFPy the user defines the mapping between the database schema and the WaterML data model. After this mapping is defined, the WaterOneFlow web methods are automatically created on the top of the existing database. No extra staging ODM database is required. WofPy is especially suitable for exposing the WaterOneFlow standard interface on existing systems that use large custom databases. The WofPy solution is also recommended for Linux server. The setup of WofPy requires administrator access to the database of the organization and experience with database management in order to setup the database access components.

For smaller organizations and individuals who are interested in preserving and sharing their observation or model simulation time series data in the form of a standards compliant WaterOneFlow web service, the existing software solutions including HydroServer and WofPy require a relatively high level of information technology skills (database administration, web server administration, accessing web server via remote desktop). The modern trend in internet technologies (Web 2.0) is to lower the bar for users who want to publish and share their content on the web (O'Reilly, 2007). This approach has been successfully applied in a wide range of domains (Haklay, 2008), including live sharing and editing of documents (Google Docs), geographic information and maps (ArcGIS Online, OpenStreetMap), personal profiles (Facebook), general knowledge (Wikipedia) and software (Codeplex). The goal of this research is to explore the design of an easily accessible Web 2.0 style data hosting application for hosting user-generated hydrologic data content and exposing this data as a WaterOneFlow web service.

SOFTWARE DESIGN REQUIREMENTS

The design requirements of the data hosting application are:

1. It should run on any free or commercial ASP.NET webhosting site

2. Any of the following databases can be used: SQL Server, SQL Server Express, MySQL, SQLite
3. The application can be deployed by one-click: just uploading required files to the web hosting folder
4. The application should provide WaterOneFlow web service version 1.1 with valid content in WaterML 1.1.
5. A simple web form is provided for uploading and editing time series data
6. A simple web form with a map interface is provided for adding or editing location (observation site) metadata

SOFTWARE ARCHITECTURE

The core component of the data hosting system is a lightweight HydroServer (HydroServer Lite) which re-uses the WaterOneFlow component of the original HydroServer software stack. A high level schema is shown in Figure 1. As seen in Figure 1 some of the advanced features of HydroServer such as the streaming data loader are not present in HydroServer Lite because these features would require administrator access and running the application under full trust. The unique features of HydroServer Lite are:

- A web service for uploading time series data. This service has the *UploadDataSeries* web method. Using this method and the POST protocol, an authorized user can send a time series in WaterML format to the server. The reason for creating a web method is to facilitate the development of 3rd party data collection application which can upload data by simply calling the UploadDataSeries web method. The web method has an *OverwriteOption* parameter. This parameter decides if values should be appended to an existing series, or if the series should be overwritten.
- An interactive web map user interface for adding a site or editing site information. Many users do not remember the coordinates of the observation site but they can locate the site in a web map by remembering nearby landmarks. The web map interface is also used for managing the existing collection of sites.
- A wide range of supported time series file formats. Three methods for entering time series data are supported: (1) Manually entering the information using a web form interface, (2) uploading the information from a local file, (3) Linking to a Google spreadsheet. The advantage of the last option is that users can maintain their data in a Google spreadsheet document. Whenever the document is updated, the time series information is automatically transferred to the data hosting system.

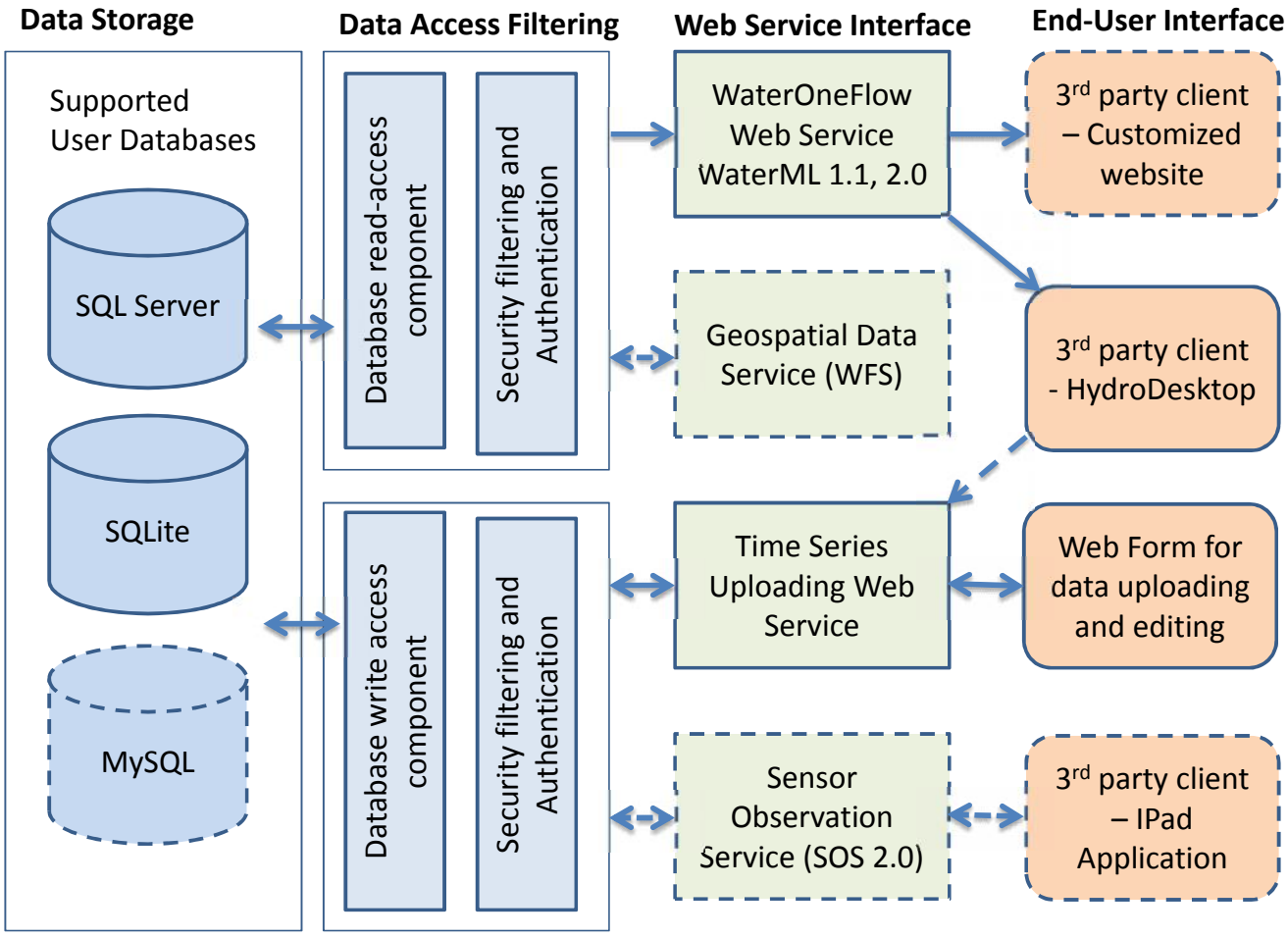

Figure 1 General architecture of HydroServer Lite and its interaction with other Hydrologic information system web services and end user tools

If customization is required, HydroServer Lite can be deployed as a standalone package on any free or commercial web hosting provider that provides ASP.NET internet web hosting services. Examples of free providers include *ASPOne* or *AppHarbor*. However, we expect that most users will take advantage of the hydrologic data hosting website. This platform is still under development but the following section describes the functional and technical design principles of the data hosting website:

The data hosting system supports multiple *User Accounts*. A user can set up a new account or sign in to an existing account. An existing identity provider (Windows Live ID, Google, Yahoo, Facebook or OpenID) can be used for the account creation and signup. Creating an account automatically creates a unique URI address with a WaterOneFlow service for the users. At this stage the user can share the link to the service with others. After signing in, the user is presented with a "My Sites" view. In the "My Sites" view an interactive map with following options is presented: (1) Define area of interest, (2) Add a new site, (3) Add observations, (4) Edit observations. With the "Add a new site" option, user can click on the map to locate the observation site and edit information about it. The newly added site appears in the map and in a table of sites. After the new site is added, the "Add Observations" option can be used to add an observation time series to the site. The observation time series can be entered using a web form,

uploaded from a local file or uploaded from an Internet location. The additional metadata about the observations (variable, method, quality control level) needs to be selected in this step. After uploading the file a preview is shown where the user can review and edit the data. Once the upload is confirmed, the series becomes accessible to the public through the WaterOneFlow web service. Optionally the user can mark the series as "private", restricting the access to authorized users.

If the "Edit observations" option is selected, then the user first selects a site using the map or the table of sites. After selecting a site the existing list of time series associated with this site is presented. After selecting a time series the data values can be edited using a web form. The user can also add an additional time series to the site by using the "Add observations" option. Figure 2 shows the process of adding a new time series to the user account.

By default the sites and series added to the user account are not registered in the HydroCatalog. However, any user can still access the data using HydroDesktop or HydroServer clients by selecting a "local service" search option and entering the service URL and accessing the uploaded time series data through HydroDesktop.

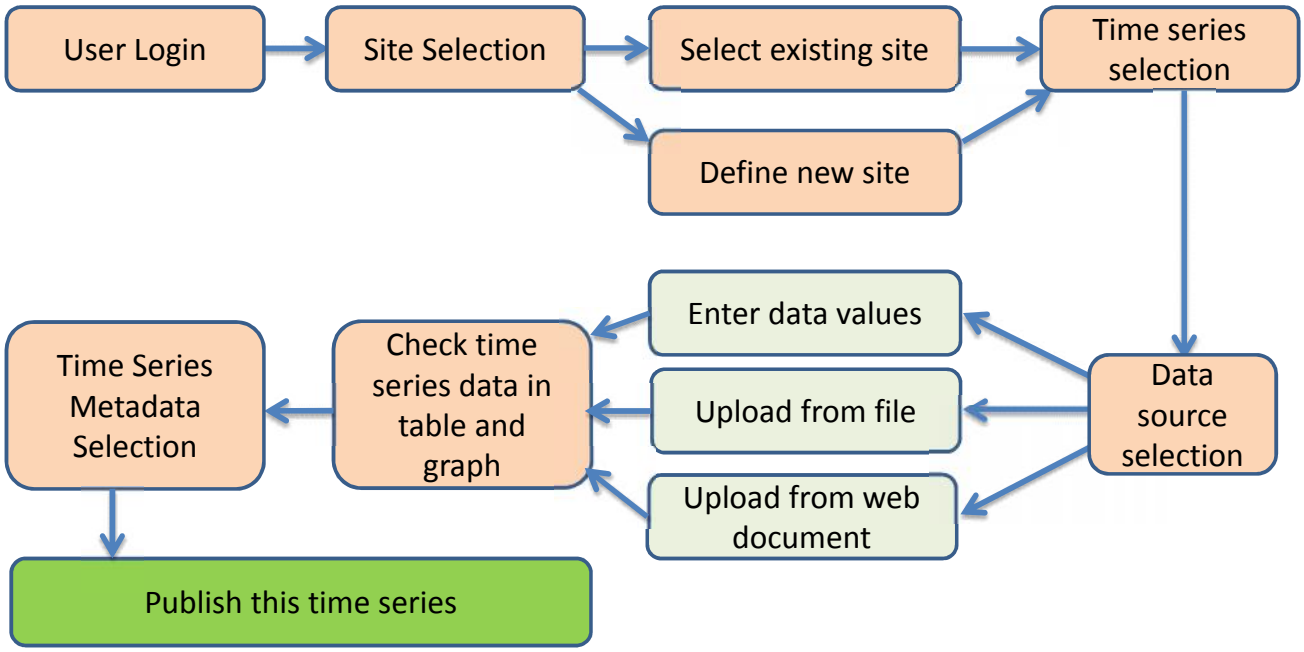

Figure 2 Flowchart - adding a new time series to the user's account on the data hosting website

CONCLUSIONS

The prototype of the lightweight HydroServer is currently being tested by several users and groups in the USA and the Czech Republic. Potential applications of the software include not only sharing field observations by small groups, but also sharing the results of hydrologic computations by students in Hydrology classes. This use case will be tested during the spring semester 2012. The current prototype of the data hosting website is limited to the ASP.NET platform. This may impose a limitation for organizations interested in deploying their own version of the data hosting website as a community portal. However, the same web services architecture that is used in the presented design may be

implemented in PHP, Java or other server side scripting languages. The complete source code of the lightweight HydroServer and data hosting website is available at: *myhydrodata.codeplex.com*. The community is invited to participate on further development of this platform. Feature requests for future releases have already been suggested by the community and include rating of user contributed data, automatic registration at HIS Central, OGC geospatial web services support and ability to create and share a user defined "theme" created by grouping user-defined selections of time series from any combination of WaterOneFlow web services.

ACKNOWLEDGMENTS

The study wouldn't be possible without the help of students from the research group of Hydrology & Climatology, Charles University Prague who volunteered to test the system prototype. Special thanks belong to the team at Utah State University who developed the full version of CUAHSI HydroServer and agreed to share the original HydroServer source code

REFERENCES

Ames, D. P. et al., 2011 (In Review). HydroDesktop: Web Services-Based Software for Hydrologic Data Discovery, Download, Visualization and Analysis. Environmental Modeling and Software.

CUAHSI, 2012. CUAHSI Hydrologic Information System http://his.cuahsi.org *Accessed* January 2012

Haklay, M., 2008. OpenStreetMap: User-Generated Street Maps. Pervasive Computing, IEEE, pp. 12-18.

Horsburgh, J., Tarboton, D. G., Maidment, D. R. & Zaslavsky, I., 2008. A relational model for environmental and water resources data. Water Resources Research.

Horsburg, J. et al., 2009. An integrated system for publishing environmental observations data. *Environmental Modelling and Software,* pp. 879-888.

Huang, M., Maidment, D. & Tian, Y., 2011. Using SOA and RIAs for water data discovery and retrieval. *Environmental Modelling & Software,* Volume 26, pp. 1309-1324.

Maidment, D., Zaslavsky, I. & Valentine, D., 2011. *CUAHSI Profile of OGC services for Time Series,* Boulder, CO: OGC Technical Committee Meeting.

O'Reilly, 2007. What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. *Communications & Strategies,* 65(1), pp. 17-37.

SDSC, 2011. *CUAHSI HIS Central Catalog and Search Web Service.* http://hiscentral.cuahsi.org. *Accessed* 1 12 2011.

Texas Water Development Board, 2012. *WOFpy – a python wrapper for WaterOneFlow services*. http://packages.python.org/WOFpy/ *Accessed* 02 01 2012.