Jun 18th, 10:40 AM - 12:00 PM

OGC and HIS: Implementing WFS and WaterML2 for HydroServer

Rohit M. Khattar
Brigham Young University, rohit@byu.edu

Dan Ames
Brigham Young University, dan.ames@byu.edu

International Environmental Modelling and Software Society (iEMSs)
7th Intl. Congress on Env. Modelling and Software, San Diego, CA, USA
Daniel P. Ames, Nigel W.T. Quinn and Andrea E. Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2014-proceedings

OGC and HIS: Implementing WFS and WaterML 2 for HydroServer

Rohit M. Khattar[a], **Dan Ames**[b]
[a] *rohit@byu.edu, Brigham Young University, Provo, Utah, USA*
[b] *dan.ames@byu.edu, Brigham Young University, Provo, Utah, USA*

Abstract: An integral step in the data life cycle (after collection and curating) is providing access and data discovery for others to explore the data. Sharing of data across various computer hardware and software platforms is enabled by tools such as the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrological Information System (HIS) which stores and manages hydrologic and climate data. This system is growing in use by water scientists globally. At the same time, the Open Geospatial Consortium (OGC) has promulgated standards used within the GIS software industry that are also based on web services technologies for data sharing. Specifically, the Web Feature Services (WFS) standard is an interface standard for allowing requests for geographical features across the web using platform-independent calls. This paper presents the design and development of a WFS that shares HIS data as geographic layers accessible by GIS applications to retrieve data from a HydroServer Lite web platform. The WFS queries data from the server and returns results conforming to another OGC standard, WaterML2. This service oriented architecture completes an essential link between two major standards in geospatial and temporal data management and sharing.

Keywords: HydroServer; CUAHSI; OGC, WaterML2; data interoperability; WFS

1 Introduction

Knowledge is power and in this digital era, knowledge has been transformed into 0's and 1's, into digital data. Data, if easily available at the right time can help mitigate disasters and save lives. Many organizations, agencies and scientists who collect hydrological and climate data realized the importance of organizing and sharing the data and hence they developed various data models suiting to their needs and type of data. As these multiple data formats began to evolve, the hydrological data network was transformed into an international air terminal where there is a wealth of data but it cannot communicate with each other due to their language barriers. Often people who collect the data and people who need the data are sitting behind different software platforms and hence data, even though available freely in a well-organized form, cannot be shared. There have been multiple standards released but so far complete interoperability has not been achieved.

One of the major organizations in the field of hydrological sciences is CUAHSI. CUAHSI is a leading effort spread across various university laboratories majorly in the United States to collect, store and share hydrological and environmental data. It is supported by the National Science Foundation (NSF) and backed up by over a 110 universities, non-profit and international affiliates, and corporate members. The CUAHSI Hydrologic Information System (HIS) is a three-part system consisting of standardized data servers (HydroServer), a centralized and searchable metadata catalog (HydroCatalog) and a desktop application for discovering and accessing data (HydroDesktop) (Ames, et al., 2012).

HydroServer is built on a standard database schema known as the Observations Data Model (ODM). ODM was designed with extensive feedback from the earth sciences community to structure data and store the associated metadata in relational tables (Horsburgh, et al.). Using the ODM database schema to store data in a HydroServer ensures that the data will be thoroughly described using metadata that is compliant with the international ISO standard for 19115 which has been adopted by many U.S. and international government agencies for documenting environmental data (e.g. see NOAA 2011).

HSL was developed as an alternative, open-source data server solution intended to be fully compliant with the CUAHSI HIS and fulfill the same functions as the full HydroServer system while being easier to install and manage and less expensive than the .NET based HydroServer. (Kadlec & Ames, 2012). HSL uses the ODM schema in its backend based on a MySQL server and PHP and HTML user interface that is designed to be intuitive and easier to use than other HIS data discovery tools such as HydroDesktop (Kadlec J. A., 2011). HSL software package includes WaterOneFlow services which were designed to serve data in WaterML 1.1 - an XML-based metadata standard for encoding hydrologic information.

The Open Geospatial Consortium (OGC) is an international voluntary consensus standards organization, originated in 1994. It comprises of more than 400 commercial, governmental, nonprofit and research organizations worldwide collaborate in a consensus process encouraging development and implementation of open standards for geospatial content and services, GIS data processing and data sharing and seeks to serve as a global forum for the collaboration of developers and users of spatial data products and services, and to advance the development of international standards for geospatial interoperability. (Open Geospatial Consortium, 2014) OGC standards are technical documents that detail interfaces or encodings. They help address data interoperability challenges and they help software developers to implement these in products or online services by two different software engineers, working independently, and the resulting components work together without further debugging.

Amongst its wide array of standards, we were interested in Web Feature Service (WFS) and WaterML. Rather than sharing geographic information at the file level using File Transfer Protocol (FTP), for example, the WFS offers direct fine-grained access to geographic information at the feature and feature property level. Most GIS applications support WFS 1.0.0 and WFS 1.1.0 servers, but are constantly being updated to incorporate the latest WFS 2.0.0 servers.

WaterML is an encoding standard that defines the information model for the representation of water observations data, with the intent of allowing the exchange of such data sets across information systems. It was developed due to combined motivation and contribution from various major hydrological organizations including U.S. Geological Survey, KISTERs AG, San Diego Supercomputer Center, CUAHSI. CUAHSI had published some existing standards such as WaterML1.0 (and 1.1), for the use in the U.S and as it gained traction in the community the name has been continued for use with the OGC standard version. OGC WaterML2.0 is, however, a harmonized information model between multiple standards from various countries with existing OGC standards. It is based on the OGC Observations and Measurements standard (ISO 19156), and therefore is significantly different from CUAHSI WaterML1.0. The most important fact was that WaterML 1.1 is not an OGC compliant encoding format.

The major issue that we address is how to make water data with both Problem: How to make water data with both geographic and temporal attributes discoverable and accessible through web services using OGC standards. Our research goal is to address this problem using WFS, WaterML, and CUAHSI WaterOneFlow web services. In this study, we propose to design and develop OGC compliant services that enable data from the CUAHSI HIS to be shared across a huge range of OGC compliant applications.

2 Methods

2.1 Software Architecture

Our software architecture brings together HIS and OGC. The services code base was built on CodeIgniter which is a PHP framework that is built on the model, view and controller (MVC) philosophy (Upton, 2007). These WFS endpoints generate geospatial site location data based on the input parameters and provide them as features to the user's GIS application. Each feature contains the data discovery endpoint for both WaterML 1.1 and WaterML2. The workflow can be represented in Figure 1.

The controllers in CodeIgniter are programmed to determine the version being requested for the WFS services and redirect the user to the appropriate service end point. As described in Figure 1, each of the service connects the MySQL backend which contains the data stored in an ODM database. For this services, the database needs to be a MySQL database, but the code can easily be reconfigured to connect to a MSSQL or PostgreSQL database. Both services return the feature set in the encoding as specified by the respective version guidelines. They link to data endpoints which are encoded either in

WaterML1.1 or WaterML2. These services once again query the data from the backend database and return the time series data in the appropriate encoding.

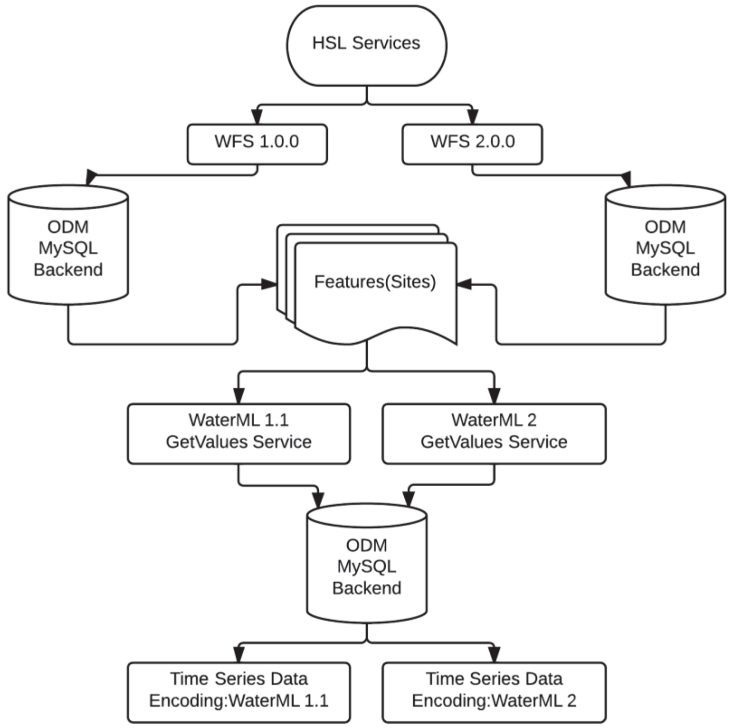

Figure 1: Services Workflow Diagram

2.2 Services Development

Each of the WFS services, baring the differences in encoding and requests, has the same basic functionality. Upon connecting to a WFS service for a specific database, it returns a list of layers from the database pertaining to the variables defined in the database that contain data points. Features for each of the layer can be requested, where each feature in that layer is a site feature with feature attributes that describe the metadata such as Latitude, Longitude, name, description and most importantly provides a URL to the WaterML2 data service for that time series. Figure 2 shows a Unified Modelling Language (UML) sequence diagram for the services.

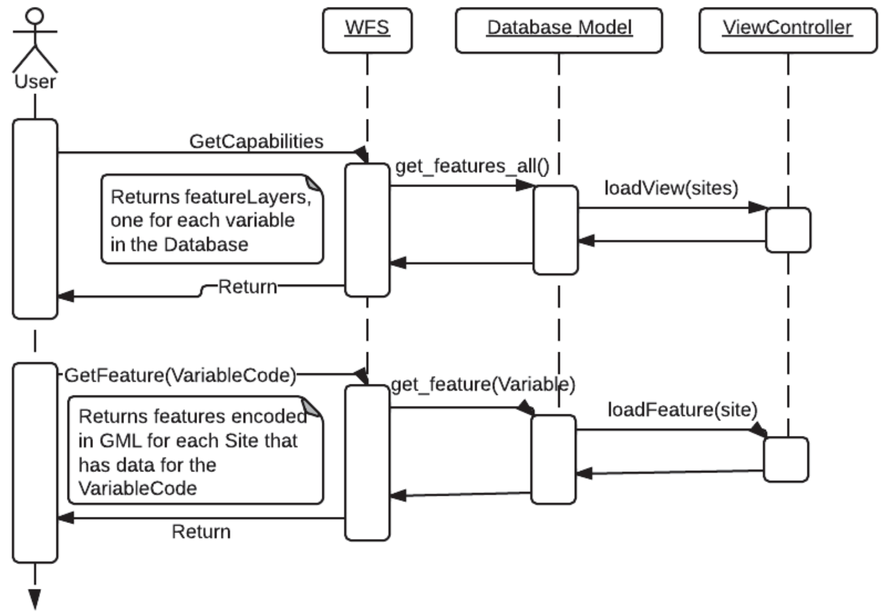

Figure 2: UML Sequence Diagram

To implement both versions of WFS, a GeoServer was installed to serve a template for developing PHP services. A GeoServer is an open source software server written in Java that allows users to share and edit geospatial data.

2.2.1 WFS 1.0.0

To implement a Basic WFS 1.0.0 service that confirms to OGC standard code: OGC 02-058, the service needs to support GET and POST operations and provide response to the following requests:

- GetCapabilities: This service returns the list of operations that one can perform on the server, provides links to the endpoints for all the requests the WFS server supports. It also returns the variables encoded as feature types that are discoverable by any GIS application. The feature types become the parameters for GetFeature requests. The bounding box in the below example is the extents of a layer containing all the sites which contain data for the specific variable.
- DescribeFeatureType: The service response contains the definition for each of the feature(Site) that will be returned from the GetFeature request. It describes its attributes and their types as well as certain restrictions on them such as if they are allowed to be Null or not.
- GetFeature: This is the final endpoint of the WFS service. It takes in a parameter (TYPENAME=VariableFeatureName) and provides the sites encoded in Geographic Markup Language (GML) 2.1.2. The encoding of the sites in GML is what makes this service compliant with OGC WFS standards. Each site contains the URLs to the WaterML services for that site and variable selection. This service was also designed to report errors in case the user is missing one of the required parameters are missing or invalid.

2.2.2 WFS 2.0.0

OGC defines the WFS 2.0.0 standards in ISO 19142 which can be obtained from their websites. This standard conformance document served as a base to develop a Basic WFS 2.0.0 service. In addition to the above mentioned services for WFS 1.0.0, it also supports the following operations:

- ListStoredQueries : This function was designed to provide a list of stored queries on the WFS server. At present, only one query, GetFeatureByID, which returns the Site as a feature object based on its ID, is supported.
- DescribeStoredQueries: Provides a description for the queries stored on the server.

WFS 2.0.0 standard also required us to update GetFeature responses to be encoded in GML 3.2. The Extensive Markup Language (XML) tags for the response differ slightly, but more or less the response is similar to WFS 1.0.0

2.2.3 Time series data in WaterML 2

WaterML2 is defined by OGC in their document: 10-126r4. The document describes the details and the XML tags that are to be employed to encode the data. It provides many categories of time series data out of which the Measurement Time Series was the best match for HIS time series data.

In addition to the above mentioned document, CUAHSI had published a mapping document that maps HIS vocabulary to WaterML2 (Valentine, 2012), which was used to code the scripts for generating WaterML2 outputs.

The scripts have support to publish multiple time series within the same document if found. The service takes the following input parameters:

- Site Code
- Variable Code
- Begin Date : Optional
- End Date : Optional

2.3 Services Evaluation

In order to be accepted as a satisfactory service to serve hydrological data, we established two constraints as the base requirement. It should confirm to OGC standards and it should be accessible through major GIS applications. To do so, a proprietary application (ArcMap 10.2) and an open source application (QGIS 2.0) were used to cover both domains of software licenses. To test the services against the base requirements the following tasks were performed:

- Testing WFS 1.0.0 with QGIS 2.0
- Testing WFS 2.0.0 with QGIS 2.0 WFS 2.0.0 plugin
- Testing WFS 1.0.0 with ArcGIS 10.2 Data Interoperability Extension
- Testing WFS 1.0.0 with WFS extension in DotSpatial Library
- Since at present there are no tools/software that consume WaterML 2 or testing suites, the data service is just made to confirm with OGC standard and HIS mapping document. (Valentine, 2012)

All the above tests were performed on a Windows 7 Enterprise operating system. To perform these tests, the service endpoint URLs were used as the input parameter and the resulting layers were added individually as well as collectively to the display. POST and GET protocols were tested too.

3 Results

The services were deployed to the databases hosted by the World Water Project at Brigham Young University, Provo.

World water project is a part of an international “grass-roots” effort to resolve the problem of data distribution and limitations in utility due to technical and administrative issues by development of tools, technologies and standards for sharing water and climate data in a manner conducive to rapid scientific development. (World Water, BYU, 2014). World water project runs an enterprise version of HSL for hosting multiple databases on the same server and is based on the concept of Software as a Service (SAAS). It was a good test case scenario as it contains a very large set of data points with diverse data from multiple sources.

3.1 Deployment to World Water

The services were deployed to the server and can be accessed simply by going to the specific database that we are looking for and appending the directory name “services” to the URL. To access the service endpoint for the test ‘Sandbox’ installation one may point the web browser to ‘http://www.worldwater.byu.edu/interactive/sandbox/services’. These services were used as our source URL’s for testing them with GIS applications.

3.2 Testing with GIS software

One of the most important tests was to confirm that data provided by the WFS services can be consumed by major GIS applications. Both applications, ArcGIS 10.2 and QGIS 2.0 support WFS 1.0.0 by default.

Another important test we performed was on the WFS extension that is built into DotSpatial. GIS applications such as HydroDesktop are built on DotSpatial as their foundation and ensuring this works makes it much easier for application developers to harvest these services.

To verify the usability of WFS 2.0.0 service, an open source plugin, WFS 2.0 Client was used. Basic WFS 2.0 needs to provide two methods to get features: by type names i.e. the variable name or by using one of the stored queries.

4 Conclusions

The goal of this study was to develop web services that will enable hydrological and climate data stored in CUAHSI HIS to be shared with various applications that support OGC standards. This was done by first building WFS endpoints within the service package for HSL. HSL, being open-source, is easily customizable to add in additional features. This solved the problem of data interoperability up to a huge extent as any software/application that is compliant with OGC will now be able to connect to a CUAHSI HIS hydroserver and have access to hydrological data in an OGC compliant WaterML2 standard. The services were tested against a variety of different GIS applications and they all connected seamlessly. Data retrieval was compliant with WaterML2 guidelines.

This study bridges the gap between data sharing and brings us closer to a unified network for hydrological data, through which we can obtain data of great value instantly without having to worry about the encoding standards or formatting.

ACKNOWLEDGMENTS

We would like to acknowledge Jiří Kadlec who helped by providing development ideas for the WFS services. Thanks to David Valentine at San Diego Computer Center for providing feedback and mapping templates for the WaterML 2 encoding.

REFERENCES

Ames, D. P., Horsburgh, J. S., Cao, Y., Kadlec, J., Whiteaker, T., & Valentine, D. (2012). HydroDesktop: Web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environmental Modelling & Software*, 146-156.

Horsburgh, J., Tarboton, D., Piasecki, M., Maidment, D., Zaslavsky, I., Valentine, D., & Whitenack, T. (n.d.). An inegrated system for publishing environmental observations data. *Environmental Modelling & Software*(24), 879-888.

Kadlec, J. A. (2011). Design and Development of Web Services for Accessing Free Hydrological Data from the Czech Republic. In J. Hřebíček, *Environmental Software Systems. Frameworks of eEnvironment.* Springer Berlin Heidelberg.

Kadlec, J., & Ames, D. (2012). Development of a lightweight HydroServer and hydrologic data hosting website. *AWRA Spring Specialty Conference on GIS and Water Resources.* New Orleans, LA.

Open Geospatial Consortium. (2014, February 2). *Vison, Mission and Goals*. Retrieved from OGC: http://www.opengeospatial.org/ogc/vision

Upton, D. (2007). *CodeIgniter for Rapid PHP Application Development.* Packt Publishing Ltd.

Valentine, D. (2012). *CUAHSI OBSERVATIONS DATA MODEL, CUAHSI WATERML 1 AND OGC WATERML 2 MAPPING.* San Diego: CUAHSI.

World Water, BYU. (2014, February 27). *Introduction*. Retrieved from World Water Project, BYU: http://www.worldwater.byu.edu