Web Services Solutions for Hydrologic Data Access and Cross-Domain Interoperability

Bora Beran
Microsoft Research
Redmond, Washington, USA
borabe@microsoft.com

David Valentine, Ilya Zaslavsky
San Diego Supercomputer Center
University of California, San Diego
La Jolla, California, USA
valentin@sdsc.edu, zaslavsk@sdsc.edu

Simon Jonathan David Cox
Commonwealth Scientific and Industrial Research Organization (CSIRO)
Bentley, Australia
simon.cox@csiro.au

John McGee
Renaissance Computing Institute
Chapel Hill, North Carolina, USA
mcgee@renci.org

***Abstract*— Agencies such as US Geological Survey (USGS), Environmental Protection Agency (EPA), National Oceanographic and Atmospheric Administration (NOAA) offer considerable amount of data on climate, hydrometry and water quality in the United States spanning from 1860s to the current day. While accessible through a web browser, data from these sources typically cannot be directly ingested by modeling or analysis tools without human intervention. Different input/output formats, syntax and terminology, and different analysis scenarios the systems were designed to support, make data discovery and retrieval a major time sink. This paper examines the web services developed as a part of Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information System (HIS) project as a means to standardize access to hydrologic data repositories, facilitate data discovery and enable direct machine-to-machine communication, and the efforts in larger scale to create a standard which is more flexible and generic yet capable of capturing the domain semantics such that interoperability with other scientific domains can be achieved losslessly.**

Keywords- Webservices; interoperability; international standards; geosciences; hydrology

I. INTRODUCTION

The world is facing major challenges associated with the environment particularly around climate change and water scarcity. Changing temperature patterns cause hydrologic cycle to become less predictable while pollution and increasing demand for water due to population growth are pushing the limits of sustainability. Coping with these issues require working across disciplines with data of varying temporal and spatial scales. For instance while flood warning systems rely on near real-time data, understanding climate change and drought patterns or making engineering decisions about structures such as dams or levees require historical data which can be in-situ point observations as well as remote sensing imagery.

In the US, Environmental Protection Agency (EPA), US Geological Survey (USGS) and National Oceanographic and Atmospheric Administration (NOAA) are the primary sources of water quality, quantity and climate data. While there are overlaps in data offerings NOAA is the main source of meteorological data, USGS stands out with its extensive water quantity (surface/subsurface) data whereas EPA focuses on environmental quality. Heterogeneity is a major issue. USGS data is available, via the National Water Information System (NWIS) in different formats including delimited text, HTML tables and USGS' own HydroML markup language. EPA is moving from delimited text to XML-based WQX (Water Quality eXchange) format. In addition to different encodings, there is no common vocabulary either. Lack of standards for hydrologic data exchange is a major problem a solution to which would eliminate the need for human involvement in data retrieval thus not only saves valuable research time but also makes it possible to implement automated workflows. This has been the main motivation behind the water data services part of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) Hydrologic Information Systems (HIS) project [1]. The project's experience in developing web for standardized access to hydrologic data sources in the United States demonstrates the challenges associated with establishing community semantics of hydrologic data exchange, formalizing the main notions of hydrologic observations, and evolution towards compliance with general data exchange protocols for cross-domain interoperability. However international aspects should also be taken into account as 145 nations have territory in the 263 trans-boundary river basins in the world and approximately one third of these basins are shared by more than two countries [2].

II. DATA COVERAGE

According to surveys, in the United States 60.8% of hydrologists in academia consider NWIS stream flow data necessary for their research [3]. NWIS is followed by NOAA's National Climatic Data Center (NCDC) precipitation data (35.1%). NCDC pan evaporation, NWIS groundwater levels, Environmental Protection Agency (EPA) Storage and Retrieval System (STORET) water quality, National Land Cover Dataset, National Elevation Dataset, State Soil Geographic (STATSGO) & Soil Survey Geographic (SSURGO) datasets, National Hydrography Dataset and remote sensing data (e.g. LANDSAT) are other datasets in the list. The CUAHSI HIS focused its attention first on the NWIS and EPA STORET as hydrologists' top two preferences with nationwide coverage and freely available data. Development of web service wrappers for hydrologic repositories at these two agencies were followed by services for Moderate Resolution Imaging Spectroradiometer (MODIS), North American Mesoscale Model (NAM) and Daily Meteorological Summaries (DAYMET) data which present gridded time series for common weather and climate variables. In addition, the hydrologic data publication workflow developed by the project, allowed other research groups, from state and local governments, academia and non-profit environmental organizations, to make their hydrologic measurements accessible through the system. The data were loaded or streamed into the CUAHSI Observations Data Model [4], exposed via the common set of web services, and registered to the CUAHSI HIS Central portal; currently over 50 community-generated data sources are published in this way.

III. HETEROGENEITY PROBLEM

Syntactic, semantic and information system disparities between web-accessible hydrologic repositories complicate their integration. To a large extent, the heterogeneities derive from the differences in the use cases envisioned in each of the agency systems, data collection and management practices, information models and internal data structures. In most cases, these characteristics are not explicitly expressed or available for review. Hence, the interoperability solutions are necessarily limited, as we attempt to capture the core semantics of hydrologic data discovery and retrieval common across different systems, and define system-specific extensions that reflect the specific intent and use cases of each agency system. Information system heterogeneity is a result of different interfaces and/or communication protocols.

Semantic heterogeneity occurs when there is no prior agreement about the meaning, interpretation or intended use of the same or related data [5]. For example equivalent measurement units can appear to be different due to several reasons such as use of different abbreviations and notations, or even typos. Table 1 gives a few examples of these differences (and errors). In the course of Water Markup Language (WaterML) 1.0 development approximately 900 units used by target repositories were reduced down to 302 common units by fixing these errors and making use of equivalences. Two mechanisms have been used within the CUAHSI HIS project to tame semantic heterogeneity. Controlled vocabularies for commonly used fields, such as units, spatial reference systems, sample medium, censor codes, etc., are managed by an online Master Controlled Vocabulary Registry available at http://his.cuahsi.org/mastercvreg/cv11.aspx and published as SOAP services, to enable vocabulary validation at the client applications. For such fields where the use of controlled vocabulary is problematic (e.g. measured parameter names), an ontology-based system is developed that lets data managers associate parameter names in their datasets with concepts in a hydrologic ontology, thus enabling semantics-based search across different repositories regardless of variable naming preferences of individual systems [6].

TABLE 1.SEMANTIC HETEROGENEITY IN MEASUREMENT UNITS

Source 1	Source 2	Note
acre feet	acre-feet	punctuation difference
micrograms per kilogram	micrograms per kilgram	spelling error
FTU	NTU	equivalent
mho	Siemens	equivalent
ppm	mg/kg	equivalent

Syntactic heterogeneity is the presence of different representations or encodings of data. Date/time formats can be given as an example where common differences are; local time vs. UTC, 12 hour clock vs. 24 hour clock and Gregorian date vs. Julian day which is common in Ameriflux data.

IV. CUAHSI WEB SERVICES

The goal of CUAHSI web services is to reconcile the aforementioned differences to the extent possible and return uniform documents regardless of the repository of origin. Hence CUAHSI HIS web services have been named WaterOneFlow; emphasizing the idea of a seamless interface through which researchers can gain access to hydrologic data from multiple heterogeneous data sources.

TABLE 2. WATERONEFLOW WEB SERVICE METHODS

Methods	**Description**
GetSiteInfo, GetSiteInfoObject	Given a site number, this method returns the site's metadata. Send the site code in this format: 'NetworkName:SiteCode'
GetSites, GetSitesObject	Given an array of site numbers, this method returns the site metadata for each one. Send the array of site codes in this format: 'NetworkName:SiteCode'
GetValues, GetValuesObject	Given a site number, a variable, a start date, and an end date, this method returns a time series. Pass in the site code and variable in this format: 'NetworkName:SiteCode' and 'NetworkName:Variable'
GetVariableInfo, GetVariableInfoObject	Given a variable code, this method returns the variable's name. Pass in the variable in this format: 'NetworkName:Variable'

WaterOneFlow follows certain rules to ensure uniformity of both input and output communication with the services. To this end web services were designed to provide output in a standard format; namely CUAHSI WaterML as part of the CUAHSI HIS project. The main purpose of WaterML has been to encode the semantics of discovery and retrieval of hydrologic time series, as commonly used by research hydrologists. This domain semantics has been derived from the CUAHSI Observations Data Model as well as from the organization, data structures and metadata exposed by several common online repositories of water quantity and water quality data. WaterML has been developed as a set of core constructs (site, variable, timeseries, etc) reflecting a common usage scenario where time series are discovered and retrieved by navigating to sites of interest and then examining parameters measured at these sites and their periods of record. As a result, WaterML offered an attractively simple formal encoding of time series exchange, which was implemented in WaterOneFlow services and field tested within a distributed system of hydrologic observatory test beds. WaterOneFlow services offer four major functions and their variants. (See Table 2) Object suffix (e.g. GetValuesObject) indicates that method returns a WaterML created by deserializing the response into an object, rather than WaterML being returned as a String. Different options are provided for users of varying levels of programming experience and not necessarily the same preferences.

```
# Data for the following site(s) are contained in
this file USGS 06090800 Missouri River at Fort
Benton MT
#
# -----------------------------------------------
# Data provided for site 06090800
#    DD parameter   Description
#    02   00060     Discharge, cubic feet per
second
#
agency_cd site_no   datetime  02_00060  02_00060_cd
5s        15s       16d        14n       10s
USGS      06090800  2009-09-06 04:00   5750      P
USGS      06090800  2009-09-06 04:15   5780      P
USGS      06090800  2009-09-06 04:30   5780      P
USGS      06090800  2009-09-06 04:45   5780      P
```

Figure 1. Sample USGS NWIS response to a data request

Figure 1 shows the output of a USGS NWIS inquiry for discharge measurements at site number 0609800 between 4:00 AM and 4:45 AM on September 6th, 2009. Figure 2 shows the response of WaterOneFlow GetValues service to the same data request. It can easily be seen that Figure 2 contains significant amount of metadata lacking in the original USGS response. Coordinates of measurement location, type of measurement (instantaneous, average, minimum, maximum, incremental etc.) and time zone are some of the additional content very important for correctly interpreting the data. This is because WaterOneFlow services are not just proxies that transform the data but are supported by a local metadata catalog or they retrieve the additional information by making several different inquiries to underlying data repositories.

WaterOneFlow services for national datasets and hydrologic observatory test-beds are operational and can be accessed at http://river.sdsc.edu/wiki/CUAHSI%20WebServices.ashx and http://www.watersnet.org/wtbs/, respectively.

There are two main deployment scenarios for WaterOneFlow services. If data is contained in CUAHSI HIS' Observations Data Model (ODM), the deployment is fairly straightforward. A different scenario is implemented when the data are housed in a remote repository such as a federal agency database accessible via a Web interface. In such cases, WaterOneFlow services can be screen scraper services aka web service wrappers. This is an error-prone approach as the services are sensitive to slight alterations of the remote web site. This bottleneck is removed as water data collection agencies develop web service interfaces to their repositories. Data repositories such as NCDC Automatic Surface Observing System (ASOS) and USGS NWIS have implemented WaterOneFlow webservices on their servers, eliminating the need for screen scraping. More repositories are expected to follow.

```xml
<timeSeries>
  <sourceInfo xsi:type="SiteInfoType">
    <siteName>Missouri River at Fort Benton MT</siteName>
    <siteCode network="NWIS">06090800</siteCode>
    <timeZoneInfo>
      <defaultTimeZone ZoneAbbreviation="MST" ZoneOffset="-07:00" />
      <daylightSavingsTimeZone ZoneAbbreviation="MDT" ZoneOffset="-06:00" />
    </timeZoneInfo>
    <geoLocation>
      <geogLocation xsi:type="LatLonPointType" srs="EPSG:4269">
        <latitude>47.81746979</latitude>
        <longitude>-110.6671586</longitude>
      </geogLocation>
    </geoLocation>
  </sourceInfo>
  <variable>
    <variableCode vocabulary="NWIS">00060</variableCode>
    <variableName>Discharge</variableName>
    <variableDescription>Discharge, cubic feet per second</variableDescription>
    <dataType>Instantaneous</dataType>
    <units unitsAbbreviation="cfs">cubic feet per second</units>
    <NoDataValue>-999999</NoDataValue>
    <timeSupport isRegular="true">
      <unit>
        <UnitName>minute</UnitName>
        <UnitType>Time</UnitType>
        <UnitAbbreviation>min</UnitAbbreviation>
      </unit>
      <timeInterval>15</timeInterval>
    </timeSupport>
  </variable>
  <values count="4">
    <value dateTime="2009-09-06T11:00:00-7:00" qualifiers="P">5750</value>
    <value dateTime="2009-09-06T11:15:00-7:00" qualifiers="P">5780</value>
    <value dateTime="2009-09-06T11:30:00-7:00" qualifiers="P">5780</value>
    <value dateTime="2009-09-06T11:45:00-7:00" qualifiers="P">5780</value>
  </values>
</timeSeries>
```


Figure 2. Excerpt from WaterOneFlow GetValues response

V. APPLICATIONS OF WATERONEFLOW SERVICES

WaterOneFlow services have been leveraged by several applications with purposes ranging from data discovery to hydrologic & water quality modeling. Macros and toolbars developed for Microsoft Excel, Matlab and ArcGIS allow importing data directly into these applications [7]. Web-based applications such as Data Access System for Hydrology (DASH) [8] and Hydroseek [6] facilitate data discovery and retrieval by providing unified map-based interfaces over multiple repositories (Figure 3).

Applications in water resources modeling make use of Open Modeling Interface (OpenMI). OpenMI defines an interface that allows time-dependent models to exchange data at run-time. Goodall et al. developed models to calculate watershed storage and water quality [9].

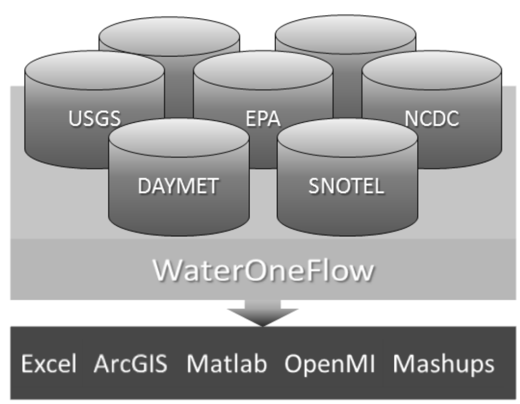

Figure 3. WaterOneFlow as a bridge to analysis and data discovery tools

Storage model is an application of conservation of mass principle and uses precipitation, streamflow (inflow-outflow) and evapotranspiration data from USGS NWIS, DAYMET and Ameriflux repositories respectively. Water quality calculations leverage USGS's SPAtially Referenced Regressions On Watershed attributes (SPARROW) model. SPARROW model performs the regression on total nitrogen loadings derived from observations of organic nitrogen, inorganic nitrogen, and flow. In this particular implementation USGS stations measuring streamflow are used along with nearby EPA stations with nitrogen concentration measurements to obtain the necessary data for the model. Once the observation results are retrieved from USGS NWIS and EPA STORET they are aligned in space and time and used as model input.

VI. INTEROPERABILITY, THE BIG PICTURE

WaterML and WaterOneFlow services have established an initial level of interoperability across hydrologic data repositories that reflected the semantics of water data discovery and retrieval common in hydrologic research. Their implementation in the context of an operational distributed system of the CUAHSI HIS project providing web service access to data measured at over 1.75 million sites in the US, allows the project team to further specify use cases and scenarios, and additional requirements for a hydrologic data exchange format. To address interoperability challenges beyond the hydrology domain, and accommodate additional usage scenarios, the approach has to be extended and harmonized with emerging standards in other domains. Several such standards are being developed under the aegis of the Open Geospatial Consortium (OGC).

A hydrology domain working group has recently been convened within OGC, to focus on formulation of interoperability requirements and scenarios in hydrology, and coordinate the development of a common exchange protocol, referred to as WaterML 2.0 operating alongside a meteorology working group under the umbrella of the OGC's Earth System Science domain working group. As part of this process WaterML is being harmonized with OGC standards for sensor/geographic data exchange to become interoperable with similar applications from different domains.

A. *Sensor Web Enablement*

Open Geospatial Consortium (OGC) provides a framework that specifies standard interfaces and encodings to facilitate exchange of geographical information. OGC's Sensor Web Enablement (SWE) initiative focuses on integration of sensors and sensor systems [10].
SWE develops standards to enable:

- Discovery of sensor systems and observations
- Determination of a sensor's capabilities
- Retrieval of sensor metadata
- Retrieval of time-series observations and coverages
- Subscription to and publishing of alerts to be issued by sensors based on certain criteria
- Tasking of sensors

The principal SWE service interface (related to the top four bullets) is called Sensor Observation Service (SOS). SOS [11] uses the OGC information standards Observations & Measurements (O&M) [12] and Sensor Model Language (SensorML) [13] for encoding observations data/metadata and sensor metadata respectively. Sensor Alert Service (SAS) and Sensor Planning Service (SPS) [14] define interfaces for subscription and tasking. SOS occupies the *services* tier shown in Figure 4. This may be compared with the IEEE 1451 family of standards which addresses the transducer interface tier.

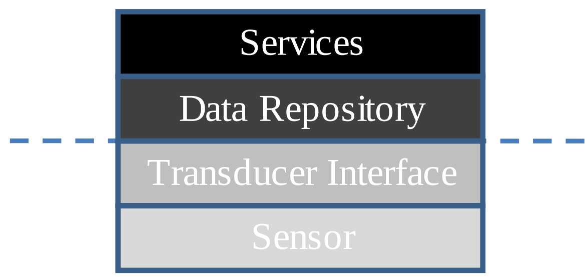

Figure 4. Generalized SWE stack

SOS defines three core and mandatory operations:

- **GetObservation** for retrieving sensor data
- **GetCapabilities** for retrieving information about the data offerings and supported functions (e.g. filters) for the service instance
- **DescribeSensor** for retrieving sensor metadata

A typical sensor data consumption scenario starts with service discovery which involves using one or more OGC Catalog Service (CS-W) [15] instances. CS-W provides an interface to a registry allowing data consumers to discover services by time period of observations, phenomena captured by observations, spatial extent, names and descriptions. Evaluating the suitability of a specific service instance utilizes the GetCapabilities operation. A GetCapabilities response contains detailed information about all of the offerings that are available from a SOS instance, which typically exposes a small constellation of sensors, details of which may be obtained through the DescribeSensor operation. GetCapabilities response also contains information on the filters supported by GetObservation operation. Filters are used to subset observation results based on temporal, spatial, logical or scalar comparison operators [11].

The SOS interface is optimized to deliver sensor-generated observations, where an observation is defined as an act that uses a procedure to determine the value of a property or phenomenon related to a feature of interest. SOS is a generic interface to observation data from any discipline. Observation semantics are provided by the definition of the *feature of interest*, the *observed property*, and the *procedure* used in generating observation results. These must be defined in the context of a particular application domain, maintained separately from the generic interface definition. The procedure may involve a sensor or observer, analytical procedure, simulation or other numerical process [12, 13].

TABLE 3. COMPARISON OF WATERONEFLOW AND SOS METHODS

Sensor Observation Service	**WaterOneFlow**	**Comments**
GetCapabilities	GetSites, GetSiteInfo	Site IDs (known as 'feature of interest' in SWE) are included in the GetCapabilities response. Capabilities are identical in all WaterOneFlow instances. GetCapabilities response contains a list of offerings, analogous to list of time series returned by WaterOneFlow.
DescribeSensor	GetVariableInfo	WaterOneFlow does not provide access to sensor or procedure descriptions. However some sensor properties are provided as part of the description of the observed variable.
GetObservation	GetValues	-
GetFeatureOfInterest	GetSiteInfo	-
DescribeObservationType	GetVariableInfo	-
DescribeFeatureType	-	Since SOS is generic, there is a specific operation to get a description of the subject of the observations. Whereas WaterML 1.0 has observation site as the only feature type.
GetFeatureOfInterestTime	GetSiteInfo	The time(s) that a mobile sensors observes a particular feature
GetResult	-	Lightweight access to values, with no metadata
DescribeResultModel	-	Since SOS/O&M are generic, a variety of result encodings may be used. This operation retrieves an explicit description of the encoding.

An SOS instance may be backed by a variety of data sources, which may be live sensors, but commonly is a data store which caches observation data. Such a cache may itself be updated through other SOS interface(s), but will commonly be updated through a private interface. (Early version SOS prototypes were even based on scraping HTML pages.) SOS merely provides a standardized http-hosted interface and request syntax, essentially a standard façade for less convenient data sources, which make them appear like a 'virtual XML document'.

Even though SOS is a fairly new standard, it is possible to see many implementations in different domains and parts of the world as an indicator of its potential to facilitate cross-domain interoperability. OOSTethys/OceansIE (Marine Science), Open architecture for Smart and Interoperable networks in Risk management based on In-situ Sensors (OSIRIS), Sensor Asia (Landslide warning, Drought monitoring) [16], Water Resources Observation Network (WRON) are examples from the United States, Europe, Asia and Australia respectively. In fact experiences from WRON project in South Esk River Catchment in the north-east of Tasmania contribute to WaterML 2.0 development. The WRON implementation communicates directly with the sensor, in contrast to WaterML 1.0 which was targeted primarily at data repositories.

B. *Water Observations Markup Language*

Water Observation Markup Language (WOML) is an application of OGC's Observations & Measurements (O&M) and Sensor Observation Service (SOS) standards for the hydrology domain. It was developed as a proof-of-concept, to evaluate the ability of the OGC standards to match the scope of WaterML v1.0.

O&M [12] decouples the generic model for an observation (with an 'observed property', 'feature of interest' 'procedure', and 'result') from the domain-specific semantics (e.g. the definition of 'stream' or 'watershed'). The latter must be provided by a separate schema, specialized for the application domain. However, recognizing that spatial sampling strategies are common across the natural sciences, O&M Part 2 [17] provides standard sampling feature types such as 'sampling point', 'sampling curve', 'specimen', which correspond to stations, profiles, transects, wells, sample etc. WOML also differs from WaterML in using controlled vocabularies from external authorities, in preference to local definitions. For example, the O&M XML implementation is a Geography Markup Language (GML) application [18], within which the Unified Code for Units of Measure (UCUM) codes [19] is recommended (when suitable). Hence WOML uses UCUM codes to scale measurement results. GML provides standard patterns for the use of URIs to link to external resources, to enable and encourage the use of pre-existing externally governed vocabularies. In this way both data structures and key aspects of the content are standardized, which leads to improved interoperability. So overall WOML is composed from O&M Parts 1 and 2, plus a lightweight domain model for hydrology (watercourse, storage, catchment), some standard vocabularies of units, sensors, interpolation rules, and state behavior and request metadata provided by SOS.

TABLE 4. SEMANTIC DIFFERENCES OVER SHARED CONCEPTS BETWEEN DATA FORMATS

WaterML 1.0	**NWIS**	**STORET**	**WQX**	**SOS**
Site	Site	Station	Monitoring Location	Feature
Lat-Long	Lat-Long	Lat-Long	Lat-Long	Arbitrary geometry (may be point coordinates)
Variable	Parameter	Characteristic	Characteristic Name	Observed property
Method	Parameter	Method	Method	Procedure
Series	Period of Record	-	Characteristic Summary	Offering

Through WOML work, use cases and experiences from WaterML and WaterOneFlow in turn are contributing design considerations for OGC standards under development. This is also a key benefit to OGC from the formation of the Hydrology working group. For example one of the key challenges in SOS/O&M is encoding time-series. Existing coverage encodings are mostly tailored for imagery, rather than functions with a temporal domain. The WaterML time-series encoding provides a good solution for handling this type of data.

From the point of view of CUAHSI, adopting externally governed standards leads to both benefits and obligations. The benefit of leveraging generic sensor and observation standards is (i) the potential for easier cross-domain data assimilation (important in hydrology, which clearly depends on meteorology, climate science, administrative and engineering information, and geology), (ii) more robust design, based on a broader set of requirements, and (iii) tool re-use. However, there are costs such as (i) dependency on third-party governance and maintenance arrangements for part of the language (ii) complexity due to specialization of a generic component, in contrast to directly designing for a limited use-case (iii) additional conformance constraints that may not be directly relevant to the application domain.

C. *Transition from WaterOneFlow to SOS*

WOML showed that O&M + SOS, customized with hydrology feature-types, property-types (variables or parameters) and sensors can support the functionality equivalent to WaterOneFlow. Table 3 shows how the SOS operations map to WaterOneFlow requests.

One of the principal goals of the OGC Hydrology Working Group is to develop WaterML v2, which will be based on the OGC SWE standards, but will address the detailed requirements identified for the WaterOneFlow services. Looking at Tables 3 and Table 4 it is possible to see that SOS is more generic and atomic, giving it much more flexibility and expressiveness as well as making it easier to parse. However this also makes SOS document structure more complex and less human-readable. While there are many conceptual overlaps at a more abstract level, hypernymy (super-ordinance) and hyponymy (sub-ordinance) are common semantics issues observed between different data sources, both in representations of the data (Table 4) and web service methods (Table 3). A consequence of this is the necessity to deal with much more complex mappings and requirement for wrapper services to often invoke multiple functions of the wrapped system and aggregate the results to be able to respond to a single request.

In order to make the adoption of SOS easier, an open–source SOS implementation is being developed which can be found at http://ogc.codeplex.com/. This work includes class libraries to support SOS and templates to simplify creation of SWE services for the Microsoft .NET environment. Libraries and templates are generic hence can be used outside the CUAHSI HIS framework and with databases other than ODM. However to simplify the migration for existing WaterOneFlow systems, a ready to use out-of-the-box web services/ODM database bundle is also included in the distribution. Operational services can be accessed at http://www.sensordatabus.org/Pages/SOS.aspx.

VII. CONCLUSION

To enable programmatic access to hydrometry/water quality databases in the United States, a set of web services has been developed. Standard web service functions (WaterOneFlow) and a markup language as the medium (CUAHSI WaterML) are used to provide a uniform view over multiple heterogeneous data sources and allow programs and modeling tools directly access and retrieve data from them without need to human intervention. This not only reduces the time spent for data discovery and preparation but also can be used in cases such as scientific work flows. WaterOneFlow services are planned to cover more data sources, offer more functions while WaterML is evolving to become an OGC standard. Web services are an important component in solving the interoperability puzzle by linking the data and applications together. However it is important to have a consensus on a standard otherwise, more time would be spent to make different standards work together. CUAHSI HIS now provides web services to USGS National Water Information System (NWIS), EPA Storage and Retrieval (STORET), Moderate Resolution Imaging Spectroradiometer (MODIS), North American Mesoscale Model (NAM) and Daily Meteorological Summaries (DAYMET) data. Through WaterOneFlow 40 other data sources are available including several international datasets.

To further enhance data interoperability within and beyond the hydrology domain, additional work focuses on harmonizing WaterML development with OGC SWE specifications. While this is a work in progress, WOML and open-source OGC libraries that couple the CUAHSI Observations Data Model with SOS interfaces are important steps towards creation and adoption of a more universal hydrologic data exchange protocol that will be both flexible and generic, at the same time providing intuitive encodings that are compatible with common hydrologic semantics.

Acknowledgment

This work has been partly supported by NSF under award EAR-0622374 (PI: D.R. Maidment), CSIRO and the Australian Bureau of Meteorology under IWIS Theme 1010.2 WIRADA project 1.4 "Water Data Transfer Standards".

References

[1] B. Beran, J. Goodall, D. Valentine, I. Zaslavsky and M. Piasecki, "Standardizing Access to Hydrologic Data Repositories through Web Services", Proc. International Conference on Advanced Geographic Information Systems & Web Services (GEOWS 2009), IEEE Computer Society, Los Alamitos, CA. Feb. 2009, pp 64-67.

[2] E. Koncagül, D. Donovan, J. Treves-Habar, S. Lee, I. Brugnon, K. Pittman, A. Núñez-Luna, C. Walther and A. Hassan (Eds), "Water: A shared responsibility: The United Nations World Water Development Report 2", UNESCO, Paris, 2006, pp 126. ISBN: 92-3-104006-5.

[3] D. Maidment (Ed), "CUAHSI HIS Status Report", September 15, 2005. pp 48-87.

[4] J.S. Horsburgh, D.G. Tarboton, D.R. Maidment and I. Zaslavsky, "A relational model for environmental and water resources data", Water Resources Research. vol. 44. doi:10.1029/2007WR006392.

[5] A.P. Sheth and J. Larsen, "Federated Database Systems for Managing Distributed, Heterogeneous and Autonomous Databases", ACM Computing Surveys: Special Issue on Heterogeneous Databases, 22(3) 183-236.

[6] B. Beran and M. Piasecki, "Engineering new paths to water data", Computers & Geosciences vol 35, issue 4, Apr. 2009, pp. 753-760.

[7] S.R. Brady, A.K. Sinha and L.C. Gundersen (Eds), "Geoinformatics 2006—Abstracts: U.S. Geological Survey Scientific Investigations Report 2006-5201", 2006. pp 23.

[8] J. L. Goodall, B.F. Robinson, F.M. Shatnawi and A.M. Castronova, "Linking Hydrologic Models and Data: The OpenMI Approach", (2007). Eos Trans. AGU, vol. 88(52), Fall Meet. Suppl., Abstract H13H-1682

[9] T. Whitenack, I. Zaslavsky, D. Valentine and D. Djokic, "Data Access System for Hydrology", (2007). Eos Trans. AGU, 88(52), Fall Meet. Suppl., Abstract H13H-1685.

[10] M. Botts, G. Percival, C. Reed and J. Davidson, "OGC Sensor Web Enablement: Overview and High Level Architecture", Open Geospatial Consortium Inc., OGC 06-052r2, 2006.

[11] A. Na and M. Priest, "Sensor Observation Service", Open Geospatial Consortium Inc., OGC 06-009r6, 2007.

[12] S. J. D. Cox, "Observations and Measurements, Part 1 - Observation Schema", Open Geospatial Consortium Inc., OGC 07-022r1, 2007.

[13] M. Botts, and A. Robin, "Sensor Model Language (SensorML)", Open Geospatial Consortium Inc., OGC 07-000, 2007.

[14] I. Simonis (Ed), "Sensor Planning Service Implementation Specification", Open Geospatial Consortium Inc., OGC 07-014r3.

[15] D. Nebert, A. Whiteside and P. Vretanos (Eds) "Catalogue Services Specification", Open Geospatial Consortium Inc., OGC 07-006r1.

[16] K. Honda, D. H. Nguyen and H. Shimamura, "Sensor Asia - Sensor Network and GIS for Environmental Monitoring in Asia", International Forum on Water Environmental Governance in Asia, pp.102-105, 2007.

[17] S. J. D. Cox, "Observations and Measurements, Part 2 – Sampling Features", Open Geospatial Consortium Inc., OGC 07-002r3, 2007.

[18] C. Portele (Ed), "OpenGIS Geography Markup Language (GML) Encoding Standard, v3.2.1", ISO 19136:2007, International Organization for Standardization, 2007.

[19] G. Schadow and C.J. McDonald, "The Unified Code for Units of Measure v1.7", Regenstrief Institute, Inc. 2008. <http://aurora.regenstrief.org/~ucum/ucum.html> 2009-09-20