Computers & Geosciences 37 (2011) 207–218

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Components of an environmental observatory information system

Jeffery S. Horsburgh [a,]*, David G. Tarboton [a], David R. Maidment [b], Ilya Zaslavsky [c]

[a] Utah Water Research Laboratory, Department of Civil and Environmental Engineering, Utah State University, Logan, UT, USA
[b] Center for Research in Water Resources, University of Texas at Austin, Austin, TX, USA
[c] San Diego Supercomputer Center, University of California at San Diego, San Diego, CA, USA

A R T I C L E I N F O

Article history:
Received 30 January 2009
Received in revised form
18 June 2010
Accepted 24 July 2010
Available online 3 November 2010

Keywords:
Cyberinfrastructure
Hydrology
Water resources
Sensor
Data model
Web service

A B S T R A C T

Recently, an initiative within the hydrologic science and environmental engineering communities has emerged for the establishment of cooperative, large-scale environmental observatories. Scientists' ability to access and use data collected within observatories to address broad research questions depends on the successful implementation of cyberinfrastructure. In this paper, we describe the architecture and functional requirements for an environmental observatory information system that supports collection, organization, storage, analysis, and publication of hydrologic observations. We then describe a unique system that has been developed to meet these requirements and that has been implemented within the Little Bear River, Utah environmental observatory test bed, as well as across a nation-wide network of 11 similar observatory test bed sites. The components demonstrated comprise an observatory information system that enables not only the management, analysis, and synthesis of environmental observations data for a single observatory, but also publication of the data on the Internet in simple to use formats that are easily accessible, discoverable by others, and interoperable with data from other observatories.

1. Introduction

Our current knowledge of the physical, chemical, and biological mechanisms controlling water quantity and quality is limited by lack of observations at the spatial density and temporal frequency needed to infer the controlling processes (Montgomery et al., 2007). Many have suggested that advancing the science of hydrology will require creation of new data networks and field experiments specifically designed to recognize the spatial and temporal heterogeneity of hydrologic processes (Kirchner, 2006; Hart and Martinez, 2006). This knowledge that current hydrologic understanding is constrained by a lack of observations at appropriate scales has motivated an initiative within the hydrologic science and environmental engineering communities towards the establishment of a network of large-scale environmental observatories (e.g., the WATer and Environmental Research Systems (WATERS) Network).[1] Environmental observatories are instrumented sites where data are collected with sufficient spatial and temporal resolution to test hypotheses in a statistically significant way. They are aimed at creating greater understanding of the earth's water and related biogeochemical cycles, and enabling improved forecasting and management of water processes.

Observatory efforts are ongoing within several scientific disciplines. These include the National Ecological Observatory Network (NEON),[2] which is planning the deployment of networked sensors and cyberinfrastructure to gather data on compelling ecological challenges; the Long Term Ecological Network (LTER),[3] which is a network of research sites that promotes synthesis and comparative research across locations and ecosystems; EarthScope,[4] which is an earth science program to explore the structure and evolution of the North American continent and understand processes controlling earthquakes and volcanoes, and many others.

Each of these observatory efforts is facing new challenges that include creating data collection infrastructure at unprecedented scales, management of increasing volumes of data as the temporal and spatial frequency of data collection increases, and the desire of the communities sponsoring these observatories for the data to become published community resources that can be used for cross-observatory analyses. Sharing data across observatories is currently complicated by the fact that observatories are geographically distributed and run by disparate research groups. The resulting heterogeneity in both data formats and descriptions limits the ability of scientists to discover and use data from more than one observatory.

* Corresponding author. Tel.: +1 435 797 2946; fax: +1 435 797 3663.
E-mail address: jeff.horsburgh@usu.edu (J.S. Horsburgh).

[1] WATer and Environmental Research Systems Network http://www.watersnet.org.
[2] National Ecological Observatory Network (NEON) http://www.neoninc.org.
[3] Long Term Ecological Network (LTER) http://www.lternet.edu.
[4] EarthScope http://www.earthscope.org.

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2010.07.003

Within the hydrologic science community, the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) is developing cyberinfrastructure to support large-scale environmental observatories (Maidment, 2005, 2008). The CUAHSI Hydrologic Information System (HIS) project has produced a variety of technologies that are advancing the way hydrologists are storing, accessing, and analyzing environmental data. These include an Observations Data Model (ODM) that provides a persistent storage mechanism for observatory data (Horsburgh et al., 2008) and Web services that provide remote programmatic access to data (Valentine et al., 2007). More information on the CUAHSI HIS project can be found at the projects website.[5]

In this paper we present advances to and implementations of cyberinfrastructure in support of environmental observatories. A single class of water resources data is addressed—time series of observational data for discharge, water quality, and climate measured at monitoring sites located at fixed points in space. We define a general architecture and functional requirements for the major components of an observatory information system required for data collection, storage, and publication. We then present an observatory information system developed for the Little Bear River, Utah environmental observatory test bed, which demonstrates seamless linkages between field sensors and a data storage, management, and publication system. The components demonstrated are reusable and enable not only the management, analysis, and synthesis of environmental observations data for a single observatory, but also publication of the data on the Internet in simple to use formats that are easily accessible, discoverable by others, and interoperable with data from other observatories. The components of the Little Bear River information system represent a new and novel approach for supporting data collection, management, and publication of observations data within environmental observatories.

2. Problem statement

Environmental observatory initiatives require investments in both capital and in development of information technology infrastructure to manage and enable observing systems. To become community resources, observatory data must be described with metadata and must be published on the Internet in formats that are easily interpretable, can be combined with data from other observatories, and support re-use for many different analyses. While there are specific software and technologies supporting parts of the overall data collection, management, and publication needs of observatories, there is currently no comprehensive, off the shelf system that supports observatory needs. Significant challenges remain in better defining the functionality of observatory information systems and the specific technologies or methods available for creating this functionality.

Many hydrologic sensor manufacturers and environmental monitoring consultants now offer for purchase services that range from sensor network design and installation to systems for publishing sensor data on the Internet. These solutions can be quite functional, but they can also be proprietary, specific to the manufacturer or consultant, and are not always designed to interoperate with other systems, making it difficult to integrate data collected using systems from multiple manufacturers. Additionally, most hydrologic data requires significant quality control processing before they can be used in scientific analyses (Mourad and Bertrand-Krajewski, 2002). There is currently little standardization in the software or methods used by scientists for performing quality control of sensor data. For example, Sciuto et al. (2009) review several methods for performing quality control of daily rainfall data, just one of many hydrologic variables measured using in-situ sensors. Guidelines for processing sensor datasets are emerging (e.g., Wagner et al., 2006), but processing data for quality control purposes can be difficult when they consist of thousands or even tens or hundreds of thousands of observations, requiring better tools that are integrated with the rest of an observatory information system for managing data and performing data quality control.

According to Beran and Piasecki (2009), the biggest challenge in seamlessly integrating multiple data sources is resolving heterogeneity issues. Many scientific investigators have devised their own systems for sharing hydrologic data, and many government agencies also have data publication systems (e.g., the United States Geological Survey's National Water Information System, the United Sates Environmental protection Agency's Storage and Retrieval System, and many others). Each of these systems has its own data storage mechanism and schema, and there is no standardization in the formats or vocabularies of data downloaded from these systems, making data integration for scientific analyses difficult (Maidment, 2008). Use of data for scientific analyses by individuals beyond those who collected the data, which would be the case for cross-observatory analyses, requires that they be adequately described and that semantic and syntactic differences in the data be resolved. Consequently, scientific data interpretation is greatly aided when data from different observational environments are published in a comparable form so that they can readily be acquired and synthesized.

There are common generic data exchange standards being developed through the World Wide Web Consortium (e.g., XML, WSDL, and SOAP) and general purpose XML specifications for describing and transmitting data. These include the Geography Markup Language (GML) (Portele, 2007), Ecological Metadata Language (EML) (EML Project Members, 2009), and others. Tools that encode the semantics of hydrologic observations in common formats and that can be used by disparate research groups to publish data from multiple observatories and sources using these common formats are needed.

3. Functionality of an observatory information system

The objective of an observatory information system is to facilitate the collection, organization, storage, analysis, and publication of environmental observations data collected within an environmental observatory. We define data collection as the process or method by which hydrologic observations are made (e.g., water temperature measurements made using an in-situ sensor and recorded on a data logger). Communication is the process by which data are transmitted from one location to another (e.g., transmitting observations from a field site to a centralized server via a telemetry network). Data organization and storage is the process by which data are converted into a format that can be used to support both data publication and analysis, including annotation of data with metadata and creation of a persistent data store. Data analysis is the process by which data are inspected, modeled, and visualized with the goal of increasing understanding of hydrologic processes, and publication is the process by which data are made universally available and presented in interoperable formats that can be discovered by scientists other than those who originally collected the data.

The general architectural and procedural components required to meet these objectives are shown in Fig. 1. They include the following: (1) data observation and communication infrastructure—the sensors and telemetry systems used to collect observations; (2) data storage and metadata—the data models, database systems,

[5] CUAHSI HIS Project website http://his.cuahsi.org.

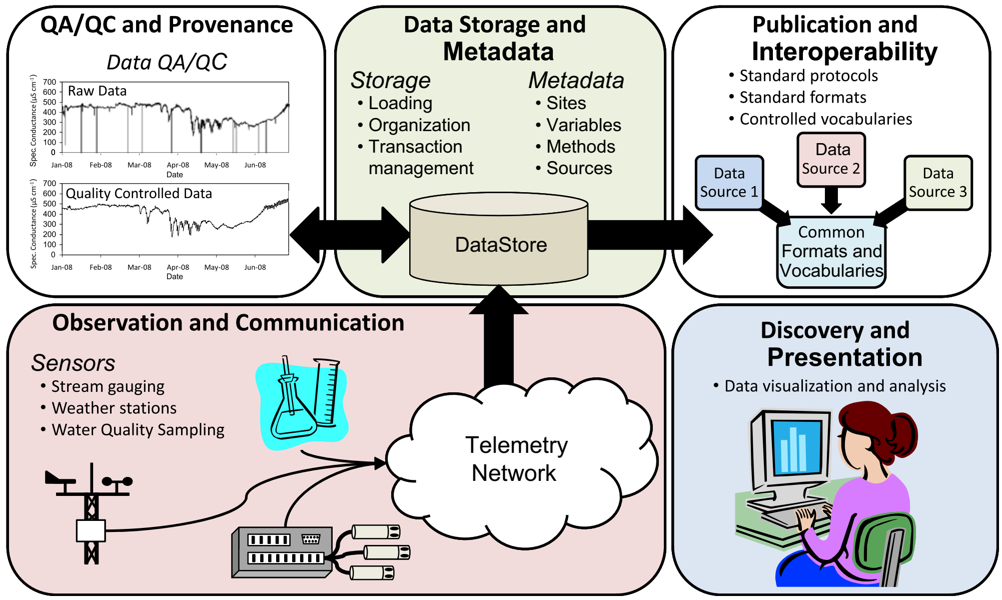

Fig. 1. General architectural components for an environmental observatory information system.

and software required for creating a persistent repository for the data; (3) quality assurance, quality control, and provenance—the software and procedures for transitioning from raw data to publishable data products; (4) publication and interoperability—the software, protocols, formats, and vocabularies used for publishing the data in interoperable formats; and (5) discovery and presentation—the tools data consumers use to get the data for the purpose of creating visualizations and analyses. The following sections describe in more detail each of these architectural pieces and present functional requirements, laying out a framework for an observatory information system.

3.1. Observation and communication infrastructure

Environmental sensors and network communications infrastructure will play a major part in proposed environmental observatories. An environmental sensor network is an array of sensor nodes and a communications system that allows their data to reach a data repository (Hart and Martinez, 2006). Variables measured at sensor nodes may include microclimate variables, precipitation chemistry variables, soil variables, stream physical and chemical variables, groundwater variables, snow variables, and many others (WATERS Network, 2008). Many of these variables are measured and reported in near real time, enabling researchers to conduct predictive modeling of water quantity and quality and enabling feedback within the monitoring systems to adjust operation and adapt monitoring in response to events (Montgomery et al., 2007).

Real time or near-real time reporting of data requires robust communications infrastructure. Currently available telemetry options include both hard wired (e.g., telephone land lines or Internet connections) and wireless solutions (e.g., cellular phone, radio, satellite). The choice of technology is dependent on the following factors: (1) required data collection and reporting frequency; (2) location and characteristics of the monitoring site; (3) power requirements and availability at remote locations; and (4) equipment and service costs. These factors present challenges for the design and implementation of sensor networks within observatories, and in current practice, communications networks may be made up of a combination of the available technologies to meet observatory data collection needs.

3.2. Data storage and metadata

Once observational data are delivered from sensor nodes to a server, they must be parsed into a persistent data store (e.g., database) so that they can be made available to data consumers. A very important part of this process is the organization of the data within the data store and matching the data with appropriate metadata. The key functionality that must be supported by the data store includes organization, storage, retrieval, and transaction management (i.e., loading, querying, and editing data). An important, value-added step involves mediation across the variety of file types and syntaxes generated by software supporting sensor and communication systems to convert data to consistent formats that can support data publication and interoperability. To date, data stores for experimental sites have ranged from file- and directory-based data structures to complex relational databases, with little coordination across sites and no accepted standard.

In many cases, environmental observations have numerical results, and, because of this, interpretation of observations requires contextual information, or metadata. Metadata is the descriptive information that explains the measurement attributes, their names, units, precision, accuracy, and data layout, as well as the lineage describing how the data was measured, acquired, or computed (Gray et al., 2005). The importance of recording metadata to help others discover and access data products is well recognized (Michener et al., 1997; Bose, 2002; Gray et al., 2005). The data store must capture not only the observation values, but their metadata as well, providing the provenance needed to trace from raw measurements to usable information and allowing observations to be unambiguously interpreted (Horsburgh et al., 2008).

3.3. Quality assurance, quality control, and data provenance

Before sensor data can be used for most applications and analyses they have to be passed through a set of quality assurance and quality control procedures (Mourad and Bertrand-Krajewski, 2002). In-situ sensors operating in harsh conditions often malfunction, many sensors are prone to drift, and data can also become corrupt when they are transmitted over communication networks. Uncorrected errors can adversely affect the value of data for scientific applications, especially if they are to be used by investigators who are not familiar with the measurement methods and conditions that may have caused the anomalies. Several studies have investigated automated anomaly detection in sensor data streams, which is particularly important in real time applications of the data and in detecting instrument malfunctions (Hill and Minsker, 2010; Hill et al., 2007; Liu et al., 2007; Mourad and Bertrand-Krajewski, 2002). Although these methods are good at detecting and flagging potentially bad sensor values, they are not always good at fixing them.

Producing high quality, continuous data streams from raw sensor output requires correcting for instrument drift, filling missing values where appropriate, and correcting spurious values. It also involves establishing linkages between raw data values and quality controlled data values to maintain data provenance. This process can be time and labor intensive, and tools that better facilitate quality assurance and quality control are needed.

3.4. Data publication and interoperability

Environmental observatories may be operated as cooperative community resources. If the data collected within environmental observatories are to become community resources, the data store and the metadata that it contains must be published in formats that enable investigators working both within and across observatories and scientific domains to easily access and interpret the data. This is especially important as it is anticipated that new discoveries will be made when data are combined or analyzed in ways that original researchers did not anticipate. Currently, all but a handful of scientific disciplines lack the technical, institutional, and cultural frameworks required to support open data access (Nature Publishing Group, 2009).

Some of the biggest challenges in achieving this are the distributed nature of the data (because there will be multiple observatories in different geographic locations) and heterogeneity within the data formats and vocabularies used to describe the data (Sheth and Larson, 1990; Colomb, 1997). Data publication systems used in observatories must not only transmit data to users both within and outside the observatory community, but they must do it in a way that overcomes semantic (e.g., differences in the language and terminology used by data collectors to describe observations) and syntactic (e.g., differences in how data collectors organize and encode their data) heterogeneity in datasets from different experimental sites (Horsburgh et al., 2009).

Although there are multiple potential solutions for publishing observatory data, recent developments in Web services and services-oriented architectures (SOAs) have led many cyberinfrastructure initiatives toward using Web SOAs because of the distributed nature of the data that they are supporting (Droegemeier et al., 2005, Youn et al., 2007; Maidment, 2008). Web services are applications that pass information between computers over the Internet, usually formatted using a platform independent markup language such as eXtensible Markup Language (XML) (Cerami, 2002; Goodall et al., 2008). SOAs rely on a collection of loosely coupled, self-contained services that communicate with each other through the Internet and that can be called from multiple software clients (e.g., Excel, Matlab, Visual Studio, etc.) in a standard fashion (Krafzig et al., 2005; Maidment, 2008).

Web services are an ideal technology upon which publication systems for distributed environmental observatories can be built because they can be used to publish data from distributed data sources, they support interoperability by using common interfaces, and they can transmit data in a common language agreed upon by the community of data collectors and users. Indeed, the emergence of SOAs is fueling movements within many scientific communities toward using standardized markup languages as self-describing, common data formats that can be used by data producers and data consumers. Examples include Earth Science Markup Language (ESML) (Ramachandran et al., 2004), Ecological Metadata Language (EML) (EML Project Members, 2008), Water Markup Language (WaterML) (Zaslavsky et al., 2007), and the Open Geospatial Consortium's (OGC) Observations and Measurements (O&M) (Cox, 2007).

According the National Research Council (2008), a robust cyberinfrastructure will provide common frameworks, components, modules, and interface models that can be used in multiple observatories or applications. Standardization upon common interfaces and formats within a SOA is the key. Each observatory can publish data using a common set of Web services that transmit data using a common language, and all of the underlying processing and complexity (which may be different from one observatory to the next) is hidden from data consumers. In addition, by standardizing the data transmission services and formats, others outside of the observatory community can also publish data using the same tools.

3.5. Data discovery and presentation

Scientists' ability to find and interpret available datasets will determine how or if the data are used (Horsburgh et al., 2009). In most cases, scientists want to download data and work with them in their own analysis environment. To do this, they need data discovery tools to help them find available data resources as well as screening and filtering tools to assist them in deciding which available data will be useful for their analyses. Hydrologic data stored in files and databases are not inherently searchable. Many hydrologic data providers such as the United States Geological Survey provide Web-based search interfaces for their data repositories through which data can be discovered and retrieved, but each data provider is different, and without knowledge of each individual website, the data cannot be discovered.

As an example of recent developments in discovery of hydrologic data, Beran and Piasecki (2009) describe a map-based search engine called Hydroseek[6] that was designed to provide users with a single interface to query and retrieve consistently formatted data from several national hydrologic data providers. Hydroseek maintains a catalog of metadata describing the data available from each provider. When users define a keyword-based search, Hydroseek searches its catalog and returns results from multiple data providers, essentially providing an equivalent to modern Internet browser search engines for hydrologic datasets.

Hydroseek provides map-based, point-and-click access to observational data, which is a powerful tool for providing users with data discovery capabilities. Users don't always know exactly what they are looking for, and the ability to see monitoring sites superimposed upon a map provides them with the spatial context they need to select data they are interested in. Juxtaposition of spatial data and time series of point observations also provides important spatial reference for interpreting the data. For example, knowing the land use distribution or terrain above a stream monitoring site is important in assessing nutrient and sediment concentrations.

Data presentation and screening tools are also an important part of the discovery process. Many users prefer to visualize datasets so that they have a better understanding of the quality and characteristics of the data before downloading them (Jeong et al., 2006). Tools for querying data and generating simple plots and descriptive statistics are generally adequate for this purpose and can also be useful for users that do not have the expertise to extract data, load it into analysis software, and then develop visualizations or analyses. By providing tools that manipulate the data automatically and do not require specialized software expertise, an observatory information system can extend the reach of the data to less technical users.

[6] Hydroseek search engine http://www.hydroseek.net/search/.

Table 1
Little Bear River monitoring sites.

Site number	Site name	Latitude	Longitude	Site description
1	Upper South Fork	41.4954	−111.818	Unregulated watershed relatively unimpacted by agricultural or urban pollutant sources
2	Lower South Fork	41.5065	−111.8151	Unregulated. Located on the South Fork below the confluence with its major tributary, Davenport Creek
3	East Fork	41.5292	−111.7993	Located below Porcupine Reservoir on the East Fork. During the summer irrigation season, the entire East Fork is diverted at this location, leaving the downstream river channel dry during most years
4	Confluence	41.5361	−111.8305	Located below the confluence of the East and South Forks. During summer, this site is primarily South Fork water as the East Fork is entirely diverted for irrigation
5	Paradise	41.5756	−111.8552	Located a short distance upstream of Hyrum Reservoir and representative of the cumulative effects of the watershed above Hyrum Reservoir
6	Wellsville	41.6435	−111.9176	Located a short distance downstream of Hyrum Reservoir. Winter flow is primarily groundwater because there are no releases from Hyrum Dam. When Hyrum Reservoir fills in the spring, high flows associated with spills from the reservoir pass this site. Summer flow is essentially groundwater as releases from Hyrum Dam are diverted for irrigation immediately below the dam and do not contribute to river flow
7	Mendon	41.7185	−111.9464	Near the terminus of the river, just upstream of the confluence with Cutler Reservoir. Influenced primarily by releases from Hyrum Reservoir and agriculture return flows
8	Lower Watershed Weather Station	41.667	−111.8906	Located near the border of the watershed and characteristic of the lower watershed below Hyrum Reservoir
9	Upper Watershed Weather Station	41.5355	−111.8059	Located near the confluence of the South and East Forks and characteristic of the mid to upper watershed

4. The Little Bear River environmental observatory test bed: a case study

As part of the planning process for a network of large-scale environmental observatories, a network of 11 observatory test bed projects[7] was created in 2006 to demonstrate technologies that could be used in the design of a national network of large-scale environmental observatories. Within each test bed, data were collected to test different scientific hypotheses related to hydrology and water resources, and investigators at each of the test beds participated with the CUAHSI HIS Team in the development and deployment of common HIS capability for publishing the observations data from all of the test beds. The goal was to create a national network of consistent data and to enable cross-domain analysis within test beds as well as cross-test bed sharing and analysis of data.

The Little Bear River of northern Utah, USA, was established as one of the test beds to test the scientific hypothesis that high-frequency discharge and water quality sensor data collected at multiple sites using in-situ sensors can improve estimates of the timing and magnitude of water quality constituent fluxes. The large volume of data needed to support this research required development of an observatory information system for collecting, organizing, managing, analyzing, and publishing the data. Where possible, existing components of the CUAHSI HIS were adopted. In addition, we developed new tools that extended the capability of the CUAHSI HIS to more fully address the functional requirements for an observatory information system. In the following sections we describe how this unique combination of tools has led to an observatory information system serving the Little Bear River and the other test beds.

4.1. Data collection and communication infrastructure: the Little Bear River sensor network

In order to generate the necessary data to enable the investigation of the hypothesis listed above, a sensor network was established that includes seven continuous stream discharge and water quality monitoring sites and 2 weather stations. At each site, a suite of sensors was connected to a Campbell Scientific, Inc. datalogger, and the data are transmitted in near real time to the Utah Water Research Laboratory (UWRL) via a telemetry network. Table 1 shows the monitoring sites and Fig. 2 shows their location within the Little Bear River watershed. Stream monitoring sites were chosen to characterize the major hydrologic and water quality regimes within the Little Bear River watershed (see Table 1). Weather station locations were chosen to characterize both the upper and lower sections of the watershed. Table 2 shows the variables measured at each type of monitoring site and the sensors used.

The telemetry network was designed to use a combination of 900 MHz spread spectrum radio links and TCP/IP Internet links to establish communications between the UWRL and each of the sites. The network enables us to monitor site status in real time and to retrieve data from each of the sites. This system was chosen because it had relatively low power requirements, it maximized the flexibility of the system for accepting new sites onto the existing network, and it did not incur monthly service costs such as those associated with cellular modem usage.

Terrain and vegetation were major challenges in the design of the radio telemetry network. Digital elevation model (DEM) based viewshed analysis using a Geographic Information System (GIS) was used to identify appropriate locations for radio repeaters so that data from the river monitoring sites, which are located at lower elevations with poor line of sight, could be transmitted to one of two remote base stations located at public schools within the watershed. Fig. 3 shows the network map for the sensor network and identifies pathways, distances, and link types between each of the remote monitoring sites and the central server located at the UWRL.

Communications with the monitoring sites are managed using Campbell Scientific's LoggerNet software.[8] LoggerNet enabled configuration of the radio linkages within the telemetry network, encoding of data collection logic into datalogger programs, and monitoring of the status of communications links within the network. The LoggerNet server was programmed to connect hourly to each remote site and download the most recent data to delimited text files, which are then stored in a location accessible on the local Intranet.

[7] WATERS Network Test Beds http://www.watersnet.org/wtbs/.

[8] Campbell Scientific http://www.campbellsci.com.

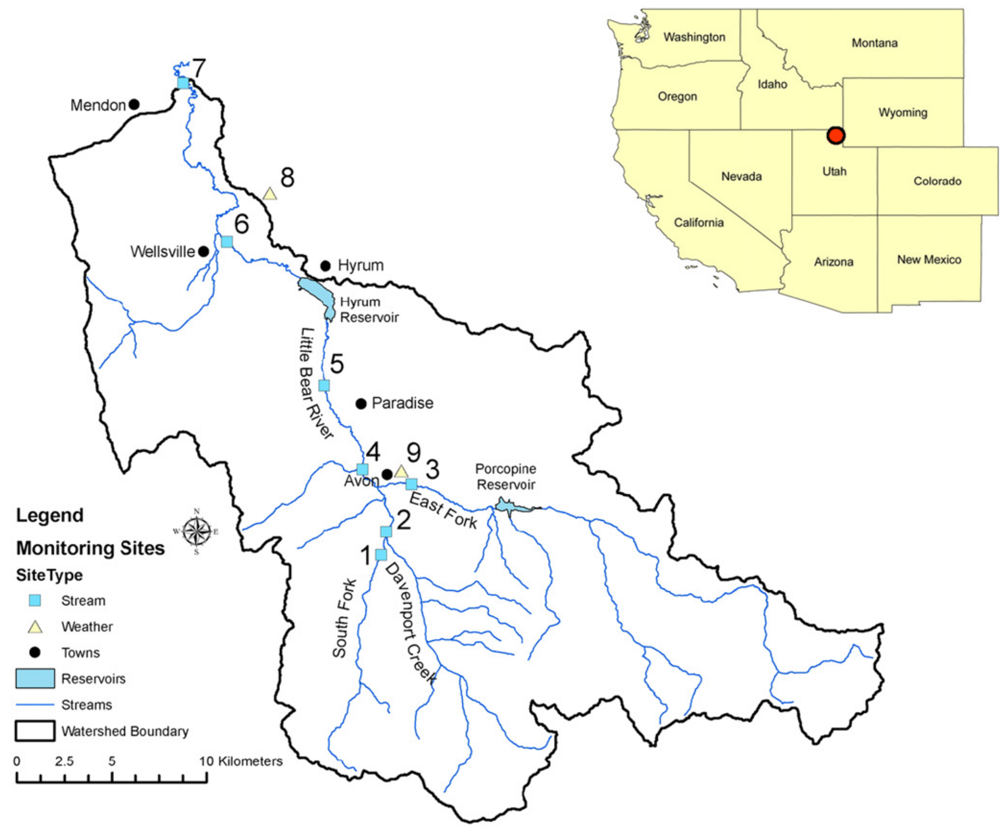

Fig. 2. Little Bear River test bed monitoring site locations.

Table 2
Sensor Specifications for Little Bear River Monitoring Sites.

Variable	Sensor	Specifications
Stream monitoring sites		
Stage	SPXD-600 Pressure Transducer KWK Technologies, Inc.	Accuracy: ±1% of the full measurement span
Turbidity	DTS-12 turbidity sensor Forest Technology Systems, Inc.	Accuracy: ±2% 0 to 500 NTU and ±4% 501 to 1600 NTU
Water temperature	Hydrolab MiniSonde5 thermistor Hach Environmental, Inc.	Accuracy: ±0.1 °C, resolution: 0.01 °C
Dissolved oxygen concentration	Hydrolab MiniSonde5 optical LDO sensor Hach Environmental, Inc.	Accuracy: ±0.1 mg L^{-1} at < 8 mg L^{-1} and ±0.2 mg L^{-1} at > 8 mg L^{-1}, resolution: 0.01 mg L^{-1}
pH	Hydrolab MiniSonde5 reference electrode Hach Environmental, Inc.	Accuracy: ±0.2 pH units, resolution: 0.01 pH units
Specific conductance	Hydrolab MiniSonde5 4-electrode, temperature compensated conductivity sensor Hach Environmental, Inc.	Accuracy: ±0.5%, resolution: 0.001 mS cm^{-1}
Weather monitoring sites		
Precipitation	TE25 tipping bucket rain gage with a 20.32 cm orifice Texas Electronics	Accuracy: ±1% up to 2.54 cm h^{-1}, resolution: 0.254 mm
Air temperature	CS215 temperature and relative humidity sensor Campbell Scientific, Inc.	Accuracy: ±0.4 °C from +5 to +40 °C, and ±0.9 °C from −40 to +70 °C
Relative humidity	CS215 temperature and relative humidity sensor Campbell Scientific, Inc.	Accuracy: ±2% at 25 °C in the 10–90% range and ±4% in the 0–100% range
Wind speed	R. M. Young Wind Sentry Set	Accuracy: ±0.5 m s^{-1}
Wind direction	R. M. Young Wind Sentry Set	Accuracy: ±0.5°
Solar radiation	PYR-P Silicon Pyranometer Apogee Instruments, Inc.	Accuracy: 5% for daily total radiation
Barometric pressure	Setra 278 Barometric Pressure Sensor	Accuracy: ±0.5 mb at +20 °C

4.2. Data storage and metadata: the Little Bear River observations database

Once the sensor data are transmitted to the UWRL, they are loaded into a database that implements the CUAHSI HIS Observations Data Model (ODM). ODM is a relational data model that was designed to be implemented within a relational database management system (RDBMS) and that defines the persistent structure of the data, including the set of attributes that accompany the data, their names, their data type, and their context (Horsburgh et al., 2008). ODM also

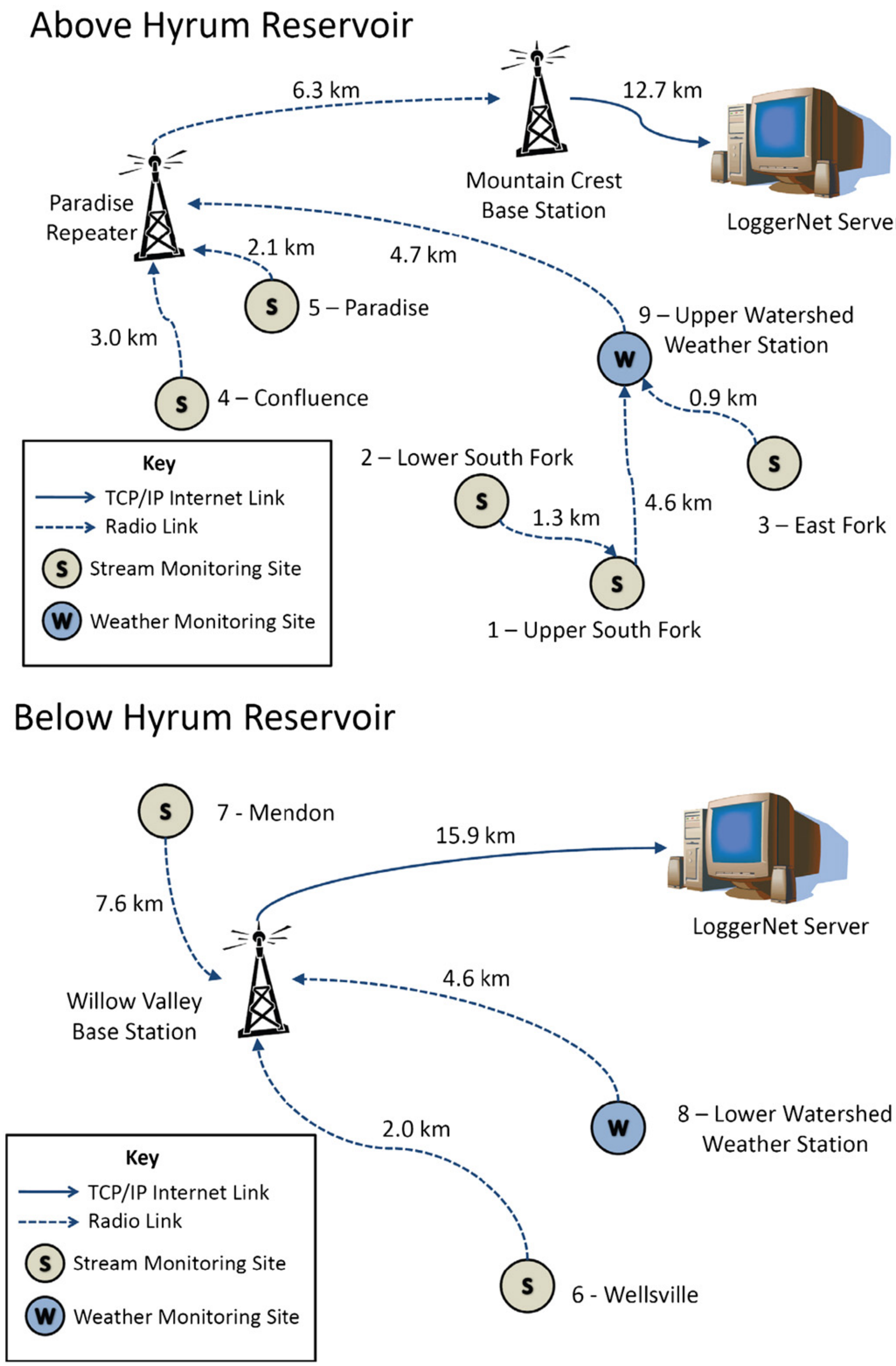

Fig. 3. Little Bear River sensor network map.

includes a set of controlled vocabularies for many of the data attributes, which are used to ensure that data stored within and across ODM instances are semantically similar (i.e., ODM database all use common attribute names and the controlled vocabularies ensure that the attribute values are the same). The Little Bear River ODM database serves as the persistent storage mechanism for the Little Bear River information system and as the active store upon which data QAQC, data publication, and data analysis are performed. ODM has been adopted by all of the test beds for storing their environmental observations data.

Each time sensor data are manually manipulated in the data loading process, there is opportunity for error. Automation is critical to avoiding human error in parsing datalogger files into the database. Because of this, we developed the ODM Streaming Data Loader (SDL) application for automating the process of loading the Little Bear River sensor data into an ODM database. The ODM

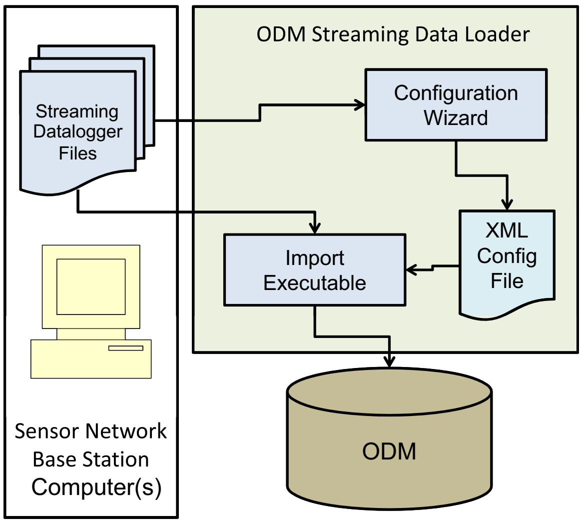

Fig. 4. Illustration of how data loading is automated using ODM Streaming Data Loader. Datalogger files are mapped to ODM using a configuration wizard, which stores mappings in an XML configuration file. An import executable then uses those mappings to parse datalogger files into an ODM database.

SDL was designed for mapping table-based datalogger files to the ODM schema and then running the data loading task periodically as new data are received. Through a wizard-based graphical user interface (GUI), the location of the datalogger file(s) on disk (or on a network shared folder or website) is defined and then all of the necessary metadata records within the ODM database are created so that the data can be loaded. Fig. 4 shows how loading of sensor data from datalogger files is automated using the ODM SDL. The ODM SDL can then be run on-demand or on a defined schedule, and, upon execution, checks each datalogger file for new observations and automatically loads them into the database.

The combination of the LoggerNet server, which manages the retrieval of data from the remote sensor nodes, and the ODM SDL, which parses the data into an ODM database, demonstrates automated integration between field sensors and a database that persistently stores the data and its metadata. Additionally, because the ODM SDL works seamlessly with the CUAHSI HIS ODM, it has now become a part of the CUAHSI HIS and is available from the CUAHSI HIS website.[9]

4.3. Quality assurance, quality control, and data provenance: ODM tools

To enable quality assurance and quality control of the sensor data, we developed software called ODM Tools that allows data managers to query, visualize, and edit data stored within an ODM database. ODM Tools provides functionality for removing obvious errors or out of range values, sensor malfunctions, and instrument drift. Users can insert, delete, adjust (by multiplying by or adding a constant value), interpolate (using adjacent values), and apply linear drift corrections to data values. Users can also annotate data values with qualifying comments, which are then stored with the data in the database.

Data editing is performed within a form that has both graphical and tabular views of the data. Fig. 5 shows the ODM Tools data editing GUI. Several data filters are available for selecting data values that may need to be edited. These include selecting data values above or below a threshold, selecting data values where gaps occur, selecting data where the change from one observation to the next is greater than some value, and selecting data occurring within a particular time interval. ODM Tools preserves the primary sensor data streams, while any edits are performed on copies derived from these data. ODM and ODM Tools preserve the provenance of the data by enabling the storage of both raw and quality controlled data within the same database and by maintaining the linkages between derived or quality controlled observations and the raw observations that they were derived from. Fig. 6 shows a portion of a time series of specific conductance, which is a measure of the electrical conductivity of water and is important in tracking dissolved water quality constituent concentrations, before and after quality control editing using ODM Tools.

4.4. Data publication and interoperability: the Little Bear River web services

The Little Bear River information system has adopted the CUAHSI HIS WaterOneFlow Web services as the main mechanism for transmitting observational data to users. The WaterOneFlow Web services were specifically developed by the CUAHSI HIS Team for publishing observations data stored in ODM databases on the Internet. The WaterOneFlow Web services respond to queries using a set of Web service methods that is the same across all implementations of the Web services, and they transmit data extracted from the Little Bear River observations database encoded using the Water Markup Language (WaterML) (Zaslavsky et al., 2007), which is an emerging standard data transmission language within the hydrologic science community. ODM and WaterML share a common information model, and there is a precise mapping between ODM and the WaterML schema. WaterOneFlow methods include *GetSites* for returning a list of sites within the database along with the metadata for each site, *GetVariableInfo* for returning a list of variables within the database along with the metadata for each variable, *GetSiteInfo* for returning a list of variables with data at a site, and *GetValues* for returning the time series of data for a site and variable combination. The Web service methods can be called from many different programming languages and software applications, including Microsoft Visual Basic, Microsoft Excel, MATLAB, and others from any computer connected to the Internet.

By adopting ODM as the storage mechanism, the WaterOneFlow Web services as the delivery mechanism, and WaterML as the format in which the data are transmitted, the Little Bear River data are published in a way that is consistent with all of the other observatory test beds (which have also adopted ODM, WaterOneFlow, and WaterML for publishing their observations data). Table 3 shows some examples of WaterOneFlow Web services published by test bed researchers, each of which is based on data stored in an ODM database. The consistent data storage format, Web services protocols, and data transmission syntax ensures that data from multiple test beds are interoperable. The XML format in which data are delivered also ensures that the data can be used regardless of platform, application, or programming language. Use of ODM as the underlying data model with its controlled vocabularies ensures that when the data from each test bed are encoded using WaterML they are consistently described and both semantic and syntactic differences are minimized.

One additional advantage to using the WaterOneFlow Web services is that high level search tools like Hydroseek (Beran and Piasecki, 2009), which is part of CUAHSI's Central HIS system and is capable of searching for data across all published WaterOneFlow Web services, and HydroDesktop (Ames et al., 2009), which is a desktop client application that expands on the search capabilities of HydroSeek by adding data visualization and analysis capabilities, can find and present data from multiple services to potential users. For example, simple keyword searches in Hydroseek and

[9] CUAHSI HIS Project website http://his.cuahsi.org.

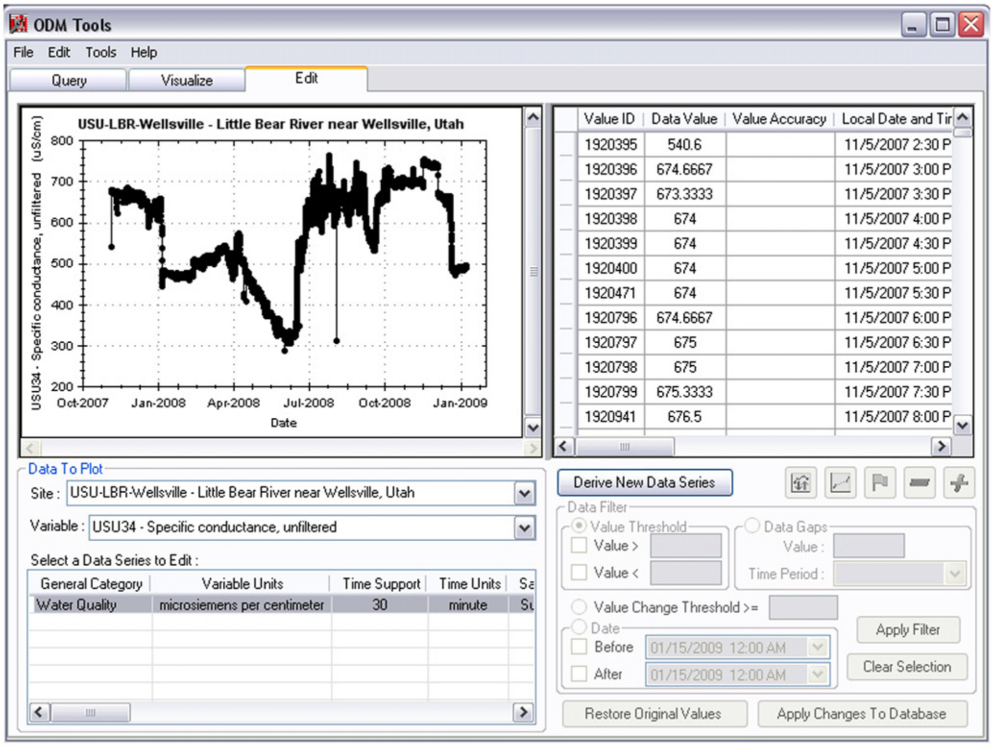

Fig. 5. ODM Tools data editing interface. A plot of a selected time series is shown (top left) and is dynamically linked to a tabular listing of data values (top right). A selection interface (bottom left) supports selection of time series of data for editing. A suite of editing tools is provided (below tabular data view), and a series of data filters for identifying potentially anomalous data values is available at bottom right.

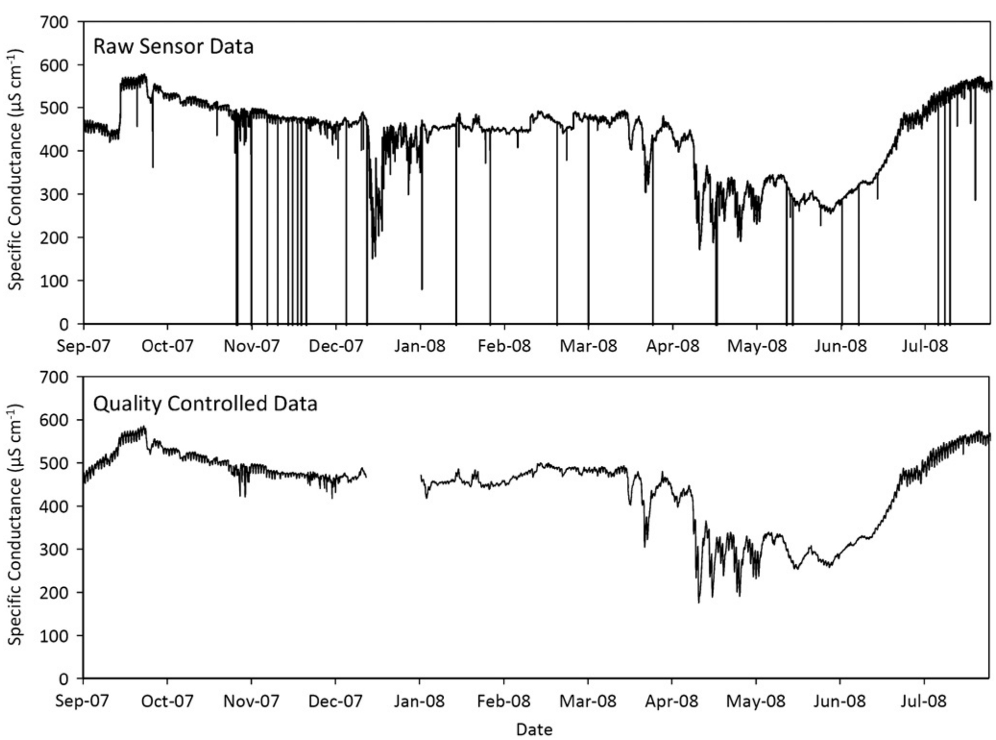

Fig. 6. Example Paradise monitoring site specific conductance data series before and after quality control editing using ODM Tools.

HydroDesktop are capable of returning observational data from each of the test beds' Web services, and, because the CUAHSI HIS Team has also built WaterOneFlow Web services for several national data providers, the same search can also return data from national data providers such as the United States Geological Survey's National Water Information System (NWIS) and the US Environmental Protection Agency's STOrage and RETrieval (STORET) system. The significance of this is not just the linkage

Table 3
Example WaterOneFlow web services published by environmental observatory test bed researchers. A full list of published web services is accessible through http://his.cuahsi.org.

Service name	Description	Service URL
Little Bear River WATERS test bed	Continuous water quality data collected in the Little Bear River, Utah test bed where researchers are investigating the use of surrogate measures such as turbidity in creating high frequency load estimates for constituents that cannot be measured continuously	http://his02.usu.edu/littlebearriver/
Crown of the Continent Observatory Snow	Data associated with the Crown of the Continent (COTC), Montana test bed, where research is focused on the study of snow-melt processes including the timing and spatial distribution of snow-melt runoff	http://his03.geol.umt.edu/COTCsnow/
Santa Fe groundwater level SRWMD	Groundwater level data from the Suwannee River Water Management District published by Santa Fe River, Florida test bed researchers who are researching the design and demonstration of a distributed sensor array for predicting water flow and nitrate flux in the Santa Fe Basin	http://ees-his06.ad.ufl.edu/SantaFe-SRGWL/
Susquehanna River Basin Hydrologic Observatory	Data associated with the Susquehanna River Basin, Pennsylvania test bed where researchers are attempting to demonstrate how a unification of modeling, existing digital data, and new data collection strategies will advance understanding of river basin water resources and support the design of hydrologic observatories	http://cbe.cae.drexel.edu/SRBHOS/
IIHR Nexrad	NEXRAD radar data from the Clear Creek, Iowa test bed, where researchers are working to establish a cyberinfrastructure-enabled, ecohydrological observatory for investigating fundamentals of runoff-driven processes as well as providing sound science for the decision-making process	http://his08.iihr.uiowa.edu/nexrad/
Baltimore waters test bed tipping bucket rain gage data	Precipitation data from the Baltimore, Maryland test bed, where researchers are evaluating the effect of subsurface infrastructure on groundwater flowpaths and fluxes, closing the urban water budget at multiple scales, and improving estimates of nutrient export from urban watersheds through a better understanding of the groundwater component of the hydrologic cycle	http://his09.umbc.edu/BaltPrecip/

with Hydroseek and HydroDesktop, but that through the adoption of a common SOA, any application developer can now program against any of the test bed or national Web services as if the data that they present were located on their own machine.

4.5. Data discovery and presentation: the Little Bear River website, map viewer, and time series analyst

A website[10] was developed for the Little Bear River that provides information about the ongoing research and links to several software applications that present the test bed data. Included is a listing of monitoring sites along with photographs, site descriptions, and information about the variables being measured and monitoring equipment installed at each one. Links are provided to launch the location of each site in a Google Maps interface. Also included in the website is a page that lists the current conditions at each site within the watershed. The current conditions page shows the latest observation of each variable at each site and is invaluable in quickly determining the status of the monitoring and telemetry system.

In addition to the website, a light weight map viewer was developed that plots the locations of the monitoring sites. It enables simple queries by allowing users to select a variable from a drop down list, which then redraws the map showing only monitoring sites with data for the selected variable. It was implemented using Google Maps and so benefits from the Google Maps base map data and Application Programmer Interface (API)[11] that enables customization of the mapping components.

When a user clicks on a monitoring site in the Little Bear River map viewer, a balloon pops up that provides information about the selected site. The balloon also provides a hyperlink to the Time Series Analyst, which is a simple, Internet-based interface to the Little Bear River observations database. Users can select a site, variable, and date range and then generate a variety of plots and summary statistics for the selected data series directly in their Web browser. They can also save the plots as images and download the data used to generate the plots. The Little Bear River map viewer and Time Series Analyst applications are available at the Little Bear River test bed website.

[10] Little Bear River test bed website http://littlebearriver.usu.edu.

[11] Google Maps Application Programmer Interface http://code.google.com/apis/maps/.

For performance purposes, both of these applications were designed to use a direct SQL connection to an ODM database. However, they were also developed to be reusable—i.e., they can be connected to multiple ODM databases. Each one has a simple query interface that allows query parameters to be passed to the application through the URL string. This is useful for launching the application in a specific state (e.g., launching the Time Series Analyst from the map viewer with a monitoring site pre-selected based on which site the user clicked on in the map). Because of its reusable design, the Time Series Analyst has now become part of the CUAHSI HIS.

Fig. 7 shows the resulting architecture of the Little Bear River observatory information system. It illustrates how users can interact with the Little Bear River observations database indirectly through the WaterOneFlow Web services, through high level search applications like HydroDesktop, and through specific tools that we have built for data discovery and presentation, including the Little Bear River map viewer and Time Series Analyst. The flexibility of this system can appeal to a broad range of users, from programmers that want to call the Web services to get data for scientific analyses to more casual users that simply want to examine a plot of the data on the Internet.

5. Conclusions

Collection, management, and publication of high frequency data present challenges for the community of scientists working toward the establishment of large-scale environmental observatories. In this paper, we have presented the architecture and functional requirements for an observatory information system that enables collecting, organizing, storing, analyzing, and publishing point observations data. The Little Bear River information system is made up of hardware and software components that together demonstrate a specific implementation of the general architecture and contribute to the cyberinfrastructure available for observatories. The unique set of tools that make up the Little Bear River information system has enabled the storage and management of all of our test bed data and open and free distribution of the data via Internet-based tools that ensure our data is available on the Internet in simple to use formats that are easily accessible, discoverable by others, and interoperable with data from the other observatory test beds.

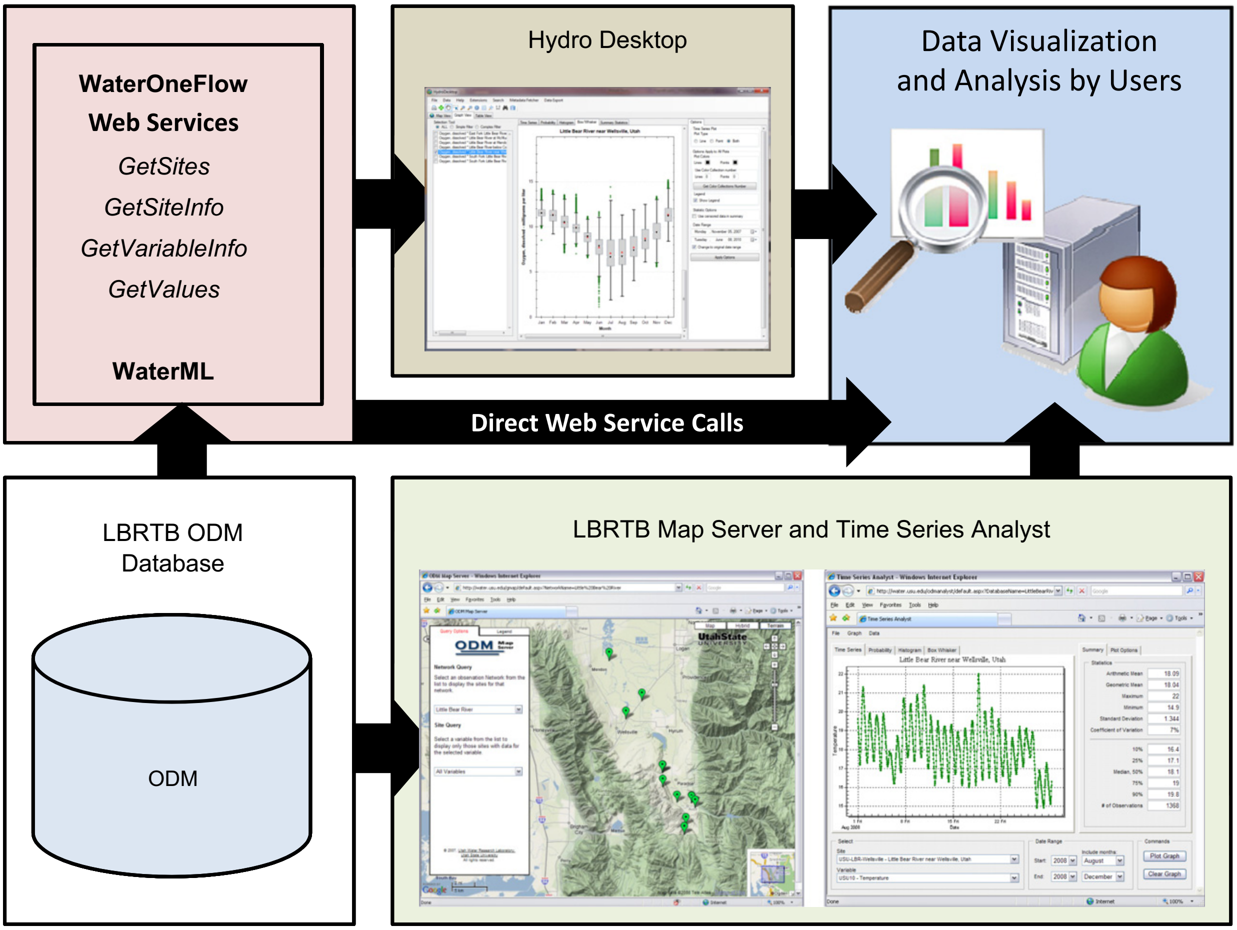

Fig. 7. Little Bear River observatory information system data discovery, visualization, and analysis components.

The components of the Little Bear River information system and CUAHSI HIS, including ODM, ODM SDL, ODM Tools, the WaterOneFlow Web services, and the Time Series Analyst, are also transferrable, meaning that anyone can use them to publish data resources. This has been demonstrated by their use at other sites within and outside of the network of observatory test beds. These tools and documentation describing how to implement them are freely available for download, lowering the barrier for implementation by others. Their applicability may also extend beyond observatories to any data-intensive study where management and publication of observational data is required.

The use of ODM and the ODM SDL has enabled automated integration between sensors in the field and a central observations database that persistently stores the data and its metadata. Automation of the data loading task eliminates potential errors and ensures that the database always contains the most recent data. ODM Tools provides a graphical user interface for transitioning data from raw sensor streams to higher level, quality checked and derived data series that can be confidently used for scientific analyses while preserving information about how the data were derived, modified, or edited.

The WaterOneFlow Web services and WaterML serve as a data transmission mechanism that is platform, software, and programming language independent, promoting interoperability among all of the observatory test beds. WaterML ensures that semantic and syntactic differences in data retrieved from all of the test beds are minimized. Through adoption of a SOA, a national network of environmental observatory test beds with consistently published scientific data has been created, and application programmers can program against the test bed Web services as if the data were located on their own machine. This is the type of functionality that must be supported within the proposed network of large-scale environmental observatories if they are to be community resources.

Data discovery and presentation tools such as the Little Bear River map viewer and the Time Series Analyst provide data users with the ability to more easily screen available data to find datasets that they may be interested in using. The linkage of the two and their accessibility within a Web browser makes the data more user-friendly to individuals who are not familiar with the Little Bear River watershed and also extends the reach of the data to individuals that may lack the skills to successfully use the Web services.

Acknowledgments

ODM, ODM SDL, ODM Tools, the WaterOneFlow Web services, and the Time Series Analyst are part of the CUAHSI HIS software stack and are available from the CUAHSI HIS website free of charge under the open-source Berkeley Software Distribution (BSD) license. The goal of the CUAHSI HIS is to advance information system technology for hydrologic science. This work was supported by the National Science Foundation under Grants EAR 0413265, EAR 0622374, and CBET 0610075. Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Ames, D.P., Horsburgh, J., Goodall, J., Whiteaker, T., Tarboton, D., Maidment, D., 2009. Introducing the open source CUAHSI Hydrologic Information System Desktop application (HIS Desktop). In: Anderssen, R.S., Braddock, R.D., Newham, L.T.H. (Eds.), Proceedings of the 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, Cairns, Australia, pp. 4353–4359. Available from: 〈http://www.mssanz.org.au/modsim09/J4/ames.pdf〉.

Beran, B., Piasecki, M., 2009. Engineering new paths to water data. Computers and Geosciences 35, 753–760. doi:10.1016/j.cageo.2008.02.017.

Bose, R., 2002. A conceptual framework for composing and managing scientific data lineage. In: Proceedings of the 14th International Conference on Scientific and Statistical Database Management. IEEE Press, Pascataway, New Jersey, pp. 15–19.

Cerami, E., 2002. Web Services Essentials: Distributed Applications with XML-RPC, SOAP, UDDI & WSDL. O'Reilly Media, Sebastopol, CA, USA, 304 pp.

Colomb, R.L., 1997. Impact of semantic heterogeneity on federating databases. Computer Journal 40 (5), 235–244. doi:10.1093/comjnl/40.5.235.

Cox, S. (Ed.), 2007. Observations and Measurements – Part1 – Observation schema. OGC07-022r1. Open Geospatial Consortium, Inc., 85 pp. Available from: 〈http://portal.opengeospatial.org/files/?artifact_id=22466〉 (accessed 20.09.10).

Droegemeier, K.K., Brewster, K., Xue, M., Weber, D., Gannon, D., Plale, B., Reed, D., Ramakrishnan, L., Alameda, J., Wilhelmson, R., Baltzer, T., Domenico, B., Murray, D., Ramamurthy, M., Wilson, A., Clark, R., Yalda, S., Graves, S., Ramachandran, R., Rushing, J., Joseph, E., 2005. Service-oriented environments for dynamically interacting with mesoscale weather. Computing in Science and Engineering 7 (6), 12–29. doi:10.1109/MCSE.2005.124.

EML Project Members, 2009. Ecological Metadata Language (EML). Available from: 〈http://knb.ecoinformatics.org/software/eml/〉 (accessed 20.09.10).

Goodall, J.L., Horsburgh, J.S., Whiteaker, T.L., Maidment, D.R., Zaslavsky, I., 2008. A first approach to Web services for the National Water Information System. Environmental Modelling and Software 23 (4), 404–411. doi:10.1016/j.envsoft.2007.01.005.

Gray, J., Liu, D.T., Nieto-Santisteban, M., Szalay, A., DeWitt, D.J., Heber, G., 2005. Scientific data management in the coming decade. SIGMOD Record 34 (4), 34–41. doi:10.1145/1107499.1107503.

Hart, J.K., Martinez, K., 2006. Environmental sensor networks: a revolution in earth system science? Earth-Science Reviews 78, 177–191 doi:10.1016/j.earscirev.2006.05.001.

Hill, D.J., Minsker, B., 2010. Anomaly detection in streaming environmental sensor data: a data driven modeling approach. Environmental Modelling and Software 25, 1014–1022. doi:10.1016/j.envsoft.2009.08.010.

Hill, D.J., Minsker, B., Amir, E., 2007. Real-time Bayesian anomaly detection for environmental sensor data. In: Proceedings of the 32nd Congress of the International Association of Hydraulic Engineering and Research (IAHR 2007), Venice, Italy, pp. 1–10. Available from: 〈http://reason.cs.uiuc.edu/eyal/papers/Bayesian-anomaly-sensor-IAHR07.pdf〉.

Horsburgh, J.S., Tarboton, D.G., Maidment, D.R., Zaslavsky, I., 2008. A relational model for environmental and water resources data. Water Resources Research 44 (W05406), 12. doi:10.1029/2007WR006392.

Horsburgh, J.S., Tarboton, D.G., Piasecki, M., Maidment, D.R., Zaslavsky, I., Valentine, D., Whitenack, T., 2009. An integrated system for publishing environmental observations data. Environmental Modelling and Software 24, 879–888. doi:10.1016/j.envsoft.2009.01.002.

Jeong, S., Liang, Y., Liang, X., 2006. Design of an integrated data retrieval, analysis, and visualization system: application in the hydrology domain. Environmental Modelling and Software 21, 1722–1740. doi:10.1016/j.envsoft.2005.09.007.

Kirchner, J.W., 2006. Getting the right answers for the right reasons: linking measurements, analyses, and models to advance the science of hydrology. Water Resources Research 42 (W03S04), 5. doi:10.1029/2005WR004362.

Krafzig, D., Banke, K., Slama, D., 2005. Enterprise SOA Service-Oriented Architecture: Best Practices. Pearson Education, Inc., Upper Saddle River, NJ, USA, 408 pp.

Liu, Y., Minsker, B., Hill, D., 2007. Cyberinfrastructure technologies to support QA/QC and event-driven analysis of distributed sensing data. In: Proceedings of the International Workshop on Advances in Hydroinformatics, Niagara Falls, Canada, pp. 1–11. Available from: 〈http://www.ncsa.illinois.edu/~yongliu/documents/CyberinfrastructureTechQAQC.pdf〉.

Maidment, D.R. (Ed.), 2005. Hydrologic Information System, Status Report. Consortium of Universities for the Advancement of Hydrologic Science Inc., Washington, DC, 224 pp. Available from: 〈http://www.cuahsi.org/docs/HISStatusSept15.pdf〉 (accessed 20.09.10).

Maidment, D.R. (Ed.), Consortium of Universities for the Advancement of Hydrologic Science Inc., Washington, DC 92 pp. Available from: 〈http://his.cuahsi.org/documents/HISOverview.pdf〉 (accessed 20.09.10).

Michener, W.K., Brunt, J.W., Helly, J.J., Kirchner, T.B., Stafford, S.G., 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications 7 (1), 330–342, doi:10.1890/1051-0761(1997)007[0330:NMFTES]2.0.CO;2.

Montgomery, J.L., Harmon, T., Kaiser, W., Sanderson, A., Haas, C.N., Hooper, R., Minsker, B., Schnoor, J., Clesceri, N.L., Graham, W., Brezonik, P., 2007. The WATERS Network: an integrated environmental observatory network for water research. Environmental Science and Technology 41 (19), 6642–6647. doi:10.1021/es072618f.

Mourad, M., Bertrand-Krajewski, J.L., 2002. A method for automatic validation of long time series of data in urban hydrology. Water Science and Technology 45 (4–5), 263–270.

National Research Council, 2008. Integrating multiscale observations of US waters, Committee on Integrated Observations for Hydrologic and Related Sciences, Water Science and Technology Board, Division on Earth and Life Studies. The National Academies Press, Washington, DC, 198 pp.

Nature Publishing Group, 2009. Data's shameful neglect. Nature 461 (7261), 145–146. doi:10.1038/461145a.

Portele, C. (Ed.), 2007. GeographyMarkupLanguage v3.2.1. Open Geospatial Consortium, Inc. OGC 07-036, 437 pp. Available from: 〈http://portal.open-geospatial.org/files/?artifact_id=20509〉 (accessed 20.09.10).

Ramachandran, R., Graves, S., Conover, H., Moe, K., 2004. Earth Science Markup Language (ESML): a solution for scientific data-application interoperability problem. Computers and Geosciences 30, 117–124 doi:10.1016/j.cageo.2003.08.010.

Sciuto, G., Bonaccorso, B., Cancelliere, A., Rossi, G., 2009. Quality control of daily rainfall data with neural networks. Journal of Hydrology 364 (1–2), 13–22. doi:10.1016/j.jhydrol.2008.10.008.

Sheth, A.P., Larson, J.A., 1990. Federated database systems for managing distributed, heterogeneous, and autonomous databases. ACM (Association for Computing Machinery) Computing Surveys: Special Issue on Heterogeneous Databases 22 (3), 183–236. doi:10.1145/96602.96604.

Valentine, D., Zaslavsky, I., Whitenack, T., Maidment, D.R., 2007. Design and implementation of CUAHSI WATERML and WaterOneFlow Web services. In: Proceedings of the Geoinformatics 2007 Conference, San Diego, California, pp. 5-3. Available from: 〈http://gsa.confex.com/gsa/2007GE/finalprogram/abstract_122329.htm〉.

Wagner, R.J., Boulger Jr., R.W., Oblinger, C.J., Smith, B.A., 2006. Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting. US Geological Survey Techniques and Methods 1-D3, Reston, Virginia, 51 pp. Available from: 〈http://pubs.water.usgs.gov/tm1d3〉 (accessed 20.09.10).

WATERS Network, 2008. Draft science, education, and design strategy for the WATer and Environmental Research Systems Network. WATERS Network Project Office, Urbana-Champaign, 142 pp. Available from: 〈http://www.watersnet.org/docs/SEDS-20080227-draft.pdf〉 (accessed 20.09.10).

Youn, C., Baru, C., Bhatia, K., Chandra, S., Lin, K., Memon, A., Memon, G., Seber, D., 2007. GEONGrid portal: design and implementations. Concurrency and Computation: Practice and Experience 19 (12), 1597–1607. doi:10.1002/cpe.1129.

Zaslavsky, I., Valentine, D., Whiteaker, T. (Eds.), 2007. CUAHSI WaterML v0.3.0. OGC07-041r1. Open Geospatial Consortium, Inc. 76 pp. Available from: 〈http://portal.opengeospatial.org/files/?artifact_id=21743〉, (accessed 20.09.10).