

1 **Components of an Integrated Environmental Observatory Information System**

2
3
4 ¹Jeffery S. Horsburgh, Utah Water Research Laboratory, Utah State University, Logan, Utah,
5 USA

6
7 David G. Tarboton, Utah Water Research Laboratory, Utah State University, Logan, Utah, USA

8
9 David R. Maidment, Center for Research in Water Resources, University of Texas at Austin,
10 Austin, Texas, USA

11
12 Ilya Zaslavsky, San Diego Supercomputer Center, University of California at San Diego, San
13 Diego, California, USA

14
15
16 Keywords: Cyberinfrastructure, Hydrology, Water Resources, Sensor, Data Model, Web Service

¹ Corresponding author. Address: 8200 Old Main Hill, Logan, UT 84322-8200, Telephone: (435) 797-2946, Fax: (435) 797-3663, Email: jeff.horsburgh@usu.edu

17 **Abstract**

18 Recently, community initiatives have emerged for the establishment of cooperative large-
19 scale environmental observatories. Scientists' ability to access and use the data collected within
20 observatories to address broad research questions will depend on the successful implementation
21 of cyberinfrastructure. In this paper, we describe the general components of an environmental
22 observatory information system for collecting, storing, and publishing point observations data.
23 We then describe the implementation of prototypes for each of the generalized components
24 within the Little Bear River observatory test bed, as well as across a nation-wide network of 11
25 observatory test bed sites. Together, these components comprise an integrated observatory
26 information system that has enabled us to not only analyze and synthesize our data to advance
27 our understanding of the Little Bear River watershed but also manage and publish all of the
28 observational data that we are collecting on the Internet in simple to use formats that are easily
29 accessible and discoverable by others.

30 **1. Introduction**

31 Our current knowledge of the physical, chemical, and biological mechanisms controlling
32 water quantity and quality is limited by lack of observations at the spatial density and temporal
33 frequency needed to infer the controlling processes (Montgomery et al., 2007). Many have
34 suggested that new data networks and field experiments that recognize the spatial and temporal
35 heterogeneity of these processes are needed to advance the science of hydrology (Kirchner,
36 2006; Hart and Martinez, 2006). This knowledge that current hydrologic understanding is
37 constrained by a lack of observations at appropriate scales has motivated community initiatives
38 (e.g., <http://www.cuahsi.org>, <http://cleaner.ncsa.uiuc.edu>, <http://www.watersnet.org/>) towards the
39 establishment of large-scale environmental observatories. The goal is to create a network of

40 instrumented sites where data are collected with unprecedented spatial and temporal resolution,
41 aiming at creating greater understanding of the earth's water and related biogeochemical cycles
42 and enabling improved forecasting and management of water processes.

43 As the amount and complexity of data grows, which will surely be the case within
44 observatories, it becomes increasingly difficult to identify trends and relationships in the data and
45 to derive information that enhances understanding using simple query and reporting tools
46 (Connolly and Begg, 2005). Using data streams from multiple sensors or from multiple sites in a
47 single analysis becomes more difficult when they consist of thousands or even tens or hundreds
48 of thousands of observations. Thus, without tools to manage and manipulate data their utility in
49 fostering process understanding is limited.

50 Additionally, it is difficult for anyone beyond those who collected the data to use them
51 for scientific analyses if they are never published or if semantic and syntactic differences
52 preclude their use in common analyses. Recently, these questions of data availability,
53 organization, publication, visualization, and analysis have come to the forefront within many
54 scientific communities. Hydrologic Information Systems (HIS) are emerging as technology to
55 address these questions in the area of hydrology and water resources.

56 Observatory initiatives will require investments in both capital and information
57 technology infrastructure to manage and enable observing systems. Comprehensive
58 infrastructure being used to capitalize on advances in information technology has been termed
59 "cyberinfrastructure" and integrates hardware for computing, data and networks, digitally-
60 enabled sensors, observatories and experimental facilities, and an interoperable suite of software
61 and middleware services and tools (National Science Foundation, 2007). The cyberinfrastructure
62 platform for environmental observatories is expected to include high-performance computing

63 tools and intensive database management for the collection, storage, and dissemination of
64 environmental data; advanced data visualization tools; community-vetted models for system and
65 process synthesis that can be used in near-real time; and collaboration and knowledge
66 networking tools that will help multidisciplinary and geographically dispersed teams of
67 researchers to work together effectively (Montgomery et al., 2007).

68 Within the hydrologic science community, the Consortium of Universities for the
69 Advancement of Hydrologic Science, Inc. (CUAHSI) has been developing cyberinfrastructure to
70 support large-scale environmental observatories (Maidment, 2005; 2008). The CUAHSI HIS
71 project has produced a variety of technologies that are advancing the way hydrologists are
72 storing, accessing, and analyzing environmental data. These include an Observations Data
73 Model (ODM) that provides a persistent storage mechanism for observatory data (Horsburgh et
74 al., 2008a) and Web services that provide remote programmatic access to data (Valentine et al.,
75 2007). More information on the CUAHSI HIS project can be found at <http://his.cuahsi.org>.

76 Although much progress has been made by the CUAHSI HIS project, significant work
77 remains to better define the components required for integrated observatory information systems,
78 the functionality each of these components should have, and the specific technologies available
79 for implementing these components. In this paper we seek to advance the cyberinfrastructure
80 available for supporting environmental observatories. A single class of water resources data is
81 addressed – observational data measured at monitoring points located at a fixed point in space.
82 We focus on data management within observatories and the software required to establish
83 seamless linkages between field sensors, a data storage and management system, applications
84 that publish the data on the Internet, and applications that support data discovery and
85 interpretation. We describe the cyberinfrastructure components required to support this

86 functionality and the functional requirements of each. Finally, we present a case study for the
87 Little Bear River observatory test bed (LBRTB), where instances of the generalized components
88 have been developed and deployed.

89 **2. Functionality of an Integrated Observatory Information System**

90 **2.1. Data Collection and Communication Infrastructure**

91 Environmental sensors and network communications infrastructure will play a major part
92 in proposed environmental observatories. An environmental sensor network is an array of sensor
93 nodes and a communications system that allows their data to reach a server (Hart and Martinez,
94 2006). Dynamic variables measured at sensor nodes may include microclimate variables,
95 precipitation chemistry variables, soil variables, stream physical and chemical variables,
96 groundwater variables, snow variables, and many others (WATERS Network, 2008). Many of
97 these variables will be measured and reported in near real time, enabling researchers to conduct
98 predictive modeling of water quantity and quality and enabling feedback within the monitoring
99 systems to adjust operation and adapt monitoring in response to events (Montgomery et al.,
100 2008).

101 Real time or near-real time reporting of data requires robust communications
102 infrastructure. Currently available telemetry options include both hard wired (e.g., telephone
103 land lines or Internet connections) and wireless solutions (e.g., cellular phone, radio, satellite).
104 The choice of technology is dependent on the following factors: 1) required data collection and
105 reporting frequency; 2) location and characteristics of the monitoring site; 3) power requirements
106 and availability at remote locations; and 4) equipment and service costs. These factors present
107 challenges for the design and implementation of observatories, and in current practice,

108 communications networks may be made up of a combination of the available technologies to
109 overcome the challenges listed above.

110 **2.2. Persistent Data Storage**

111 Once observational data are delivered from sensor nodes to a server, they must be parsed
112 into a persistent data store (i.e., database). The key functionality that must be supported includes
113 persistent storage, retrieval, and transaction management (i.e., loading, querying, and editing
114 data). To date, this has been done in a number of ways, ranging from file- and directory-based
115 data structures to complex relational databases.

116 Environmental observations are not self describing, and, because of this, interpretation of
117 observations requires contextual information, or metadata. Metadata is the descriptive
118 information that explains the measurement attributes, their names, units, precision, accuracy, and
119 data layout, as well as the lineage describing how the data was measured, acquired, or computed
120 (Gray et al., 2005). The importance of recording metadata to help others discover and access
121 data products is well recognized (Bose, 2002; Michener et al., 1997; Gray et al., 2005). The data
122 store must capture not only the observation values, but their metadata as well, providing
123 traceable heritage from raw measurements to usable information and allowing observations to be
124 unambiguously interpreted (Horsburgh et al., 2008a).

125 **2.3. Quality Assurance, Quality Control, and Data Provenance**

126 Before sensor data can be used for most applications and analyses they have to be passed
127 through a set of quality assurance and quality control procedures (Mourad and Bertrand-
128 Krajewski, 2002). In-situ sensors operating in harsh conditions often malfunction, many sensors
129 are prone to drift, and data can also become corrupt when they are transmitted over
130 communication networks. Uncorrected errors can adversely affect the value of data for scientific

131 applications, especially if they are to be used by investigators that are not familiar with the
132 measurement methods and conditions that may have caused the anomalies. Several studies have
133 investigated automated anomaly detection in sensor data streams, which is particularly important
134 in real time applications of the data and in detecting instrument malfunctions (Hill et al., 2007;
135 Liu et al., 2007; Mourad and Bertrand-Krajewski, 2002). Although these methods are good at
136 detecting and flagging potentially bad sensor values, they are not always good at fixing them.

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

142 **2.4. Data Publication and Interoperability**

143 Environmental observatories may be operated as cooperative community resources. To
144 become so, observatory data must be published in a way that investigators working both within
145 and across observatories and scientific domains can easily access and interpret the data. One of
146 the biggest challenges in achieving this is heterogeneity within the data formats and vocabularies
147 used to describe the data (Colomb, 1997; Sheth and Larson, 1990). Data publication systems
148 used in observatories must not only transmit data to users, but they must do it in a way that
149 overcomes semantic and syntactic heterogeneity in datasets (Horsburgh et al., 2008b).

150 Web services are applications that pass information between computers over the Internet,
151 usually formatted using a platform independent markup language such as eXtensible Markup
152 Language (XML) (Goodall et al., 2008). Many cyberinfrastructure initiatives are now using
153 Web service-oriented architectures (SOA) (Maidment, 2008; Droegemeier et al., 2005, Youn et

154 al., 2007, others). SOAs rely on a collection of loosely coupled, self-contained services that
155 communicate with each other through the Internet and that can be called from multiple clients
156 (e.g., Excel, Matlab, Visual Studio, etc.) in a standard fashion (Maidment, 2008). Web services
157 can be used to accomplish data publication and interoperability by transmitting data over the
158 Internet in a platform independent format like XML.

159 The proposed network of observatories will require distributed cyberinfrastructure.
160 According the National Research Council (2006), a robust cyberinfrastructure will provide
161 common frameworks, components, modules, and interface models that can be used in multiple
162 observatories or applications. Standardization upon a SOA is the key. Each observatory can
163 publish data using a common set of web services that transmit data using a common language,
164 and all of the underlying processing and complexity (which may be different from one
165 observatory to the next) is hidden from data consumers. In addition, by standardizing the data
166 transmission services and formats, others outside of the observatory community can also publish
167 data using the same tools.

168 **2.5. Data Discovery, Visualization, and Analysis**

169 Scientists' ability to find and interpret available datasets will determine how or if the data
170 are used (Horsburgh et al., 2008b). In most cases, scientists want to download data and work
171 with them in their own analysis environment. To do this, they need screening tools to assist
172 them in deciding which data will be useful for their analyses. Map-based, point-and-click access
173 to observational data can be a powerful tool for providing users with data discovery capabilities.
174 Beran and Piasecki (2008) describe a map-based search engine called Hydroseek
175 (<http://www.hydroseek.org>) that was designed to provide users with a single interface to query
176 and retrieve consistently formatted data from several national hydrologic data providers. Users

177 don't always know exactly what they are looking for, and the ability to see monitoring sites
178 superimposed upon a map provides them with the spatial context they need to select data they are
179 interested in. Juxtaposition of spatial data and time series of point observations also provides
180 important spatial reference for interpreting the data. For example, knowing the land use
181 distribution or terrain above a stream monitoring site is important in assessing nutrient and
182 sediment concentrations.

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

190 **3. The Little Bear River Environmental Observatory Test Bed: A Case Study**

191 As part of the planning process for a network of large-scale environmental observatories,
192 a network of 11 observatory test bed projects was created in 2006 to demonstrate technologies
193 that could be used in the design of a national network of large-scale environmental observatories.
194 The test bed network has been the proving grounds for the CUAHSI HIS. Investigators at each
195 of the test beds participated with the CUAHSI HIS Team in the development and deployment of
196 common HIS capability for publishing observations data. The goal was to create a national
197 network of consistent data and to enable cross-domain analysis within test beds as well as cross-
198 test bed sharing and analysis of data. More information about the test beds can be found at the
199 following URL (<http://www.watersnet.org/wtbs/index.html>).

200 The Little Bear River of northern Utah, USA, was established as one of the test beds to
201 test the hypothesis that high-frequency discharge and water quality data collected at multiple
202 sites using in-situ sensors can improve hydrologic and hydrochemical process understanding.
203 The data intensive nature of this research required development of prototypes for the components
204 of an integrated observatory information system to provide tools for managing the data that are
205 being collected. In addition, components of the CUAHSI HIS were adopted for publishing the
206 LBRTB data in a way that is consistent with all of the other observatory test beds. In the
207 following sections we describe each of the components, the role that they have served, and how
208 the combination of these components has led to an integrated observatory information system for
209 the LBRTB.

210 **3.1. Data Collection and Communication Infrastructure: The LBRTB Sensor Network**

211 In order to generate the necessary data to enable the investigation of the hypothesis listed
212 above, a sensor network was established that includes seven continuous stream discharge and
213 water quality monitoring sites and 2 weather stations. At each site, a suite of sensors was
214 connected to a Campbell Scientific, Inc. datalogger, and the data are transmitted in near real time
215 to the Utah Water Research Laboratory (UWRL) via a telemetry network. Table 1 lists the
216 monitoring sites and Figure 1 shows their location within the Little Bear River watershed. Table
217 2 lists the variables measured at each type of monitoring site, the data collection frequency, and
218 the sensors used.

219 The telemetry network was designed to use a combination of 900 MHz spread spectrum
220 radio links and TCP/IP Internet links to establish communications between the UWRL and each
221 of the sites. The network enables us to monitor site status in real time and to retrieve data from
222 each of the sites. This system was chosen because it eliminated monthly service costs, it had

223 relatively low power requirements, and it maximized the flexibility of the system for accepting
224 new sites onto the existing network.

225 Terrain and vegetation were major challenges in the design of the radio telemetry
226 network. Digital elevation model (DEM) based watershed analysis using a Geographic
227 Information System (GIS) was used to identify appropriate locations for radio repeaters so that
228 data from the river monitoring sites, which are located at lower elevations with poor line of sight,
229 could be transmitted to one of two remote base stations located at public schools within the
230 watershed. Figure 2 shows the network map for the sensor network and identifies pathways,
231 distances, and link types between each of the remote monitoring sites and the central server
232 located at the UWRL.

233 Communications with the monitoring sites are managed using Campbell Scientific's
234 LoggerNet software (<http://www.campbellsci.com>). LoggerNet has enabled configuration of the
235 radio linkages within the telemetry network, encoding of data collection logic into datalogger
236 programs, and monitoring of the status of communications links within the network. The
237 LoggerNet server is programmed to connect hourly to each remote site and download the most
238 recent data to delimited text files, which are then stored in a location accessible on the local
239 Intranet.

240 **3.2. Persistent Data Storage: The LBRTB Observations Database**

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

246 controlled vocabularies for many of the data attributes, which are used to ensure that data stored
247 within and across ODM instances are semantically similar. The Little Bear River ODM database
248 serves as the persistent storage mechanism for the LBRTB information system.

249 Because there is opportunity for error each time sensor data are handled, automation is
250 critical to avoiding errors in parsing the datalogger files into the database. Because of this, we
251 developed the ODM Streaming Data Loader (SDL) application, which allows users to map table-
252 based datalogger files to the ODM schema and then run the data loading task periodically as new
253 data are received. Through a wizard-based graphical user interface (GUI), users define the
254 location of the datalogger file(s) on disk (or on a network shared folder or website) and then
255 create all of the necessary metadata records within the ODM database so that the data can be
256 loaded. Figure 3 shows the GUI for the ODM SDL. The ODM SDL can then be run manually
257 or on a user defined schedule, and, upon execution, checks each datalogger file for new
258 observations and automatically loads them into the database. The combination of the LoggerNet
259 server, which manages the retrieval of data from the remote sensor nodes, and the ODM SDL,
260 which parses the data into an ODM database, demonstrates automated integration between field
261 sensors and a central observations database that persistently stores the data and its metadata.

262 **3.3. Quality Assurance, Quality Control, and Data Provenance: ODM Tools**

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

269 Data editing is performed within a form that has both graphical and tabular views of the
270 data. Figure 4 shows the ODM Tools data editing interface. Several data filters are available for
271 selecting data values that may need to be edited. These include selecting data values above or
272 below a threshold, selecting data values where gaps occur, selecting data where the change from
273 one observation to the next is greater than some value, and selecting data occurring within a
274 particular time interval. ODM Tools preserves the primary sensor data streams, while any edits
275 are performed on copies derived from these data. ODM and ODM Tools preserve the
276 provenance of the data by maintaining the linkages between derived or quality controlled
277 observations and the raw observations that they were derived from. Figure 5 shows a portion of
278 a specific conductance time series before and after quality control editing using ODM Tools.

279 **3.4. Data Publication and Interoperability: The LBRTB Web Services**

280 The LBRTB information system has adopted the CUAHSI HIS WaterOneFlow (WOF)
281 web services as the main mechanism for communicating observational data to users. The WOF
282 web services respond to queries using a standard set of web service methods, and transmit data
283 extracted from the LBRTB observations database encoded using Water Markup Language
284 (WaterML) (Zaslavsky et al., 2007). WOF methods include *GetSites* for returning a list of sites
285 within the database along with the metadata for each site, *GetVariableInfo* for returning a list of
286 variables within the database along with the metadata for each variable, *GetSiteInfo* for returning
287 a list of variables with data at a site, and *GetValues* for returning the time series of data for a site
288 and variable combination. The web service methods can be called from many different
289 programming languages and software applications, including Microsoft Visual Basic, Microsoft
290 Excel, MATLAB, and others from anywhere an Internet connection is available.

291 By adopting the WOF web services and WaterML, the LBRTB data are published in a
292 format that is consistent with all of the other observatory test beds (which have also adopted the
293 WOF web services), creating a network of consistently published scientific data. WaterML
294 serves as a standard data transmission language, ensuring that data retrieved from all of the test
295 beds is syntactically similar and promoting interoperability of the data through the use of
296 standard web services protocols and an XML schema that is platform, application, and
297 programming language independent. The use of ODM as the underlying data model with its
298 controlled vocabularies ensures that when the data from each test bed are encoded using
299 WaterML they are consistently described and semantically similar.

300 One additional advantage to using the WOF web services is that high level search tools
301 like Hydroseek, which is part of CUAHSI's Central HIS system and is capable of consuming
302 WOF web services, can find and present data to potential users. Simple keyword searches in
303 Hydroseek are now capable of returning observational data from each of the test beds' web
304 services as well as from national data providers such as the United States Geological Survey and
305 the U.S. Environmental Protection Agency. The significance of this is not just the linkage with
306 Hydroseek, but that through the adoption of a common SOA, application developers can now
307 program against any of the test bed web services as if the data that they present were located on
308 their own machine.

309 **3.5. Data Discovery, Visualization, and Analysis: The LBRTB Website, Map Server,**
310 **and Time Series Analyst**

311 A website was developed for the LBRTB that provides information about the ongoing
312 research and links to several applications that present the LBRTB data
313 (<http://littlebearriver.usu.edu>). Included is a listing of monitoring sites along with photographs,
314 site descriptions, and information about the variables being measured and monitoring equipment

316 installed at each one. Links are provided to launch the location of each site in a Google Maps
317 interface. Also included in the website is a listing of the current conditions within the watershed.
318 This listing shows the latest observation of each variable at each site and is invaluable in
319 determining the status of the monitoring and telemetry system.

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

326 When a user clicks on a monitoring site in the LBRTB map server, a balloon pops up that
327 provides information about the selected site. The balloon also provides a hyperlink to the Time
328 Series Analyst, which is a simple, Internet-based interface to the LBRTB observations database.
329 Users can select a site, variable, and date range and then generate a variety of plots and summary
330 statistics for the selected data series directly in their Web browser. They can also save the plots
331 as images and download the data used to generate the plots. The LBRTB map server and Time
332 Series Analyst applications are available at the Little Bear River test bed website
333 (<http://littlebearriver.usu.edu>).

334 Both of these applications were designed to use a direct SQL connection to an ODM
335 database. However, they were also developed to be generic and reusable – i.e., they can be
336 connected to multiple ODM databases. Each one has a simple query interface that allows query
337 parameters to be passed to the application through the URL string. This is useful for launching

338 the application in a specific state (e.g., launching the Time Series Analyst from the map server
339 with a monitoring site pre-selected based on which site the user clicked on in the map).

340 Figure 6 shows the resulting architecture of the LBRTB observatory information system.
341 It illustrates how users can interact with the LBRTB observations database indirectly through the
342 WOF web services, through high level search applications like Hydroseek, and through specific
343 tools that we have built for data discovery, visualization, and analysis, including the LBRTB
344 map server and Time Series Analyst. The flexibility of this system can appeal to a broad range
345 of users, from programmers that want to call the web services to get data for scientific analyses
346 to more casual users that simply want to examine a plot of the data on the Internet.

347 **4. Discussion and Conclusions**

348 Collection and management of high frequency data present challenges for the community
349 of scientists working toward the establishment of large-scale environmental observatories. In
350 this paper, we have presented the architecture and functional requirements of an observatory
351 information system for collecting, storing, and publishing point observations data. The LBRTB
352 observatory information system is made up of hardware and software components that together
353 demonstrate a specific implementation of the general architecture and advance the
354 cyberinfrastructure available for observatories. The LBRTB information system has enabled the
355 storage and management of all of our test bed data and open and free distribution of the data via
356 simple to use, Internet-based tools. The components of the LBRTB information system are also
357 transferrable, which has been demonstrated by their use at other sites within the network of test
358 beds. Their applicability may extend beyond observatories to all data-intensive studies where
359 management and publication of observational data is required.

360 The use of ODM and the ODM SDL has enabled automated integration between sensors
361 in the field and a central observations database that persistently stores the data and its metadata.
362 Automation of the data loading task eliminates potential errors and ensures that the database
363 always contains the most recent data. ODM Tools provides graphical tools for transitioning data
364 from raw sensor streams to higher level, quality checked and derived data series that can be
365 confidently used for scientific analyses while preserving the provenance of the data through the
366 editing process.

367 The WOF web services and WaterML serve as a data communication mechanism that is
368 platform, application, and programming language independent, promoting interoperability
369 among all of the observatory test beds. WaterML ensures that data retrieved from all of the test
370 beds is syntactically and semantically similar. Through adoption of a SOA, the test beds have
371 created a national network of consistently published scientific data, and application programmers
372 can program against their web services as if the data were located on their own machine. This is
373 the type of functionality that must be supported within the proposed network of large-scale
374 environmental observatories if they are to be community resources.

375 Data discovery and visualization tools such as the LBRTB map server and the Time
376 Series Analyst provide data users with the ability to quickly screen data to find what they are
377 most interested in. The linkage of the two and their accessibility within a Web browser makes
378 the data more user-friendly to individuals who are not familiar with the Little Bear River
379 watershed and also extends the reach of the data to individuals that may lack the skills to
380 successfully use the web services.

381 **Acknowledgments**

382 ODM, ODM SDL, ODM Tools and the WOF Web services were developed as part of the
383 CUAHSI HIS project whose goal is to advance information system technology for hydrologic
384 science. These software tools are available from the CUAHSI HIS website <http://his.cuahsi.org>
385 free of charge under the open-source Berkeley Software Distribution (BSD) license. This work
386 was supported by the National Science Foundation under grants EAR 0413265, EAR 0622374,
387 and CBET 0610075. Any opinions, findings and conclusions or recommendations expressed in
388 this material are those of the authors and do not necessarily reflect the views of the National
389 Science Foundation.

390 **References**

- 391 Beran, B., Piasecki, M., 2008. Engineering new paths to water data, *Computers and Geosciences*,
392 In press. doi:10.1016/j.cageo.2008.02.017.
393
- 394 Bose, R., 2002. A conceptual framework for composing and managing scientific data lineage, in
395 *Proceedings of the 14th International Conference on Scientific and Statistical Database*
396 *Management*, pp. 15-19, IEEE Press, Piscataway, N. J.
397
- 398 Colomb, R. L., 1997. Impact of semantic heterogeneity on federating databases, *The Computer*
399 *Journal*, 40 (5), 235-244, doi:10.1093/comjnl/40.5.235.
400
- 401 Connolly, T., Begg, C., 2005. *Database Systems A Practical Approach to Design,*
402 *Implementation, and Management*, 4th edn., Addison-Wesley, Harlow, U. K.
403
- 404 Droegemeier, K. K., Brewster, K., Xue, M., Weber, D., Gannon, D., Plale, B., Reed, D.,
405 Ramakrishnan, L., Alameda, J., Wilhelmson, R., Baltzer, T., Domenico, B., Murray, D.,
406 Ramamurthy, M., Wilson, A., Clark, R., Yalda, S., Graves, S., Ramachandran, R., Rushing,
407 J., Joseph, E., 2005. Service-oriented environments for dynamically interacting with
408 mesoscale weather, *Computing in Science & Engineering*, 7 (6), 12-29,
409 doi:10.1109/MCSE.2005.124.
410
- 411 Goodall, J. L., Horsburgh, J. S., Whiteaker, T. L., Maidment, D. R., Zaslavsky, I., 2008. A first
412 approach to Web services for the National Water Information System, *Environmental*
413 *Modelling & Software*, 23 (4), 404-411, doi:10.1016/j.envsoft.2007.01.005.
414
- 415 Gray, J., Liu, D. T., Nieto-Santisteban, M., Szalay, A., DeWitt, D. J., Heber, G., 2005. Scientific
416 data management in the coming decade, *SIGMOD Record*, 34 (4), 34-41,
417 doi:10.1145/1107499.1107503.
418
- 419 Hart, J. K., Martinez, K. 2006. Environmental sensor networks: A revolution in earth system
420 science?, *Earth-Science Reviews*, 78, 177-191, doi:10.1016/j.earscirev.2006.05.001.
421
- 422 Hill, D. J., Minsker, B., Amir, E., 2007. Real-time Bayesian anomaly detection for
423 environmental sensor data, In: *Proceedings: 32nd Congress of the International Association of*
424 *Hydraulic Engineering and Research (IAHR 2007)*, Venice, Italy. (Available at
425 <http://reason.cs.uiuc.edu/eyal/papers/Bayesian-anomaly-sensor-IAHR07.pdf>)
426
- 427 Horsburgh, J. S., Tarboton, D. G., Maidment, D. R., Zaslavsky, I., 2008a. A relational model for
428 environmental and water resources data, *Water Resources Research*, 44, W05406,
429 doi:10.1029/2007WR006392.
430
- 431 Horsburgh, J. S., Tarboton, D. G., Piasecki, M., Maidment, D. R., Zaslavsky, I., Valentine, D.,
432 Whitenack, T., 2008b. An integrated system for publishing environmental observations data,
433 *Environmental Modelling & Software*, doi:10.1016/j.envsoft.2009.01.002 (In Press)
434

- 435 Jeong, S., Liang, Y., Liang, X., 2006. Design of an integrated data retrieval, analysis, and
 436 visualization system: Application in the hydrology domain, *Environmental Modelling &*
 437 *Software*, 21, 1722-1740, doi:10.1016/j.envsoft.2005.09.007.
 438
- 439 Kirchner, J. W., 2006. Getting the right answers for the right reasons: Linking measurements,
 440 analyses, and models to advance the science of hydrology, *Water Resources Research*, 42,
 441 W03S04, doi:10.1029/2005WR004362.
 442
- 443 Liu, Y., Minsker, B., Hill, D., 2007. Cyberinfrastructure technologies to support QA/QC and
 444 event-driven analysis of distributed sensing data, In: *Proceedings: International Workshop*
 445 *on Advances in Hydroinformatics*, Niagara Falls, Canada. (Available at
 446 [http://colab.ncsa.uiuc.edu/EESHG1/Documents/repository/conference/CyberinfrastructureTe](http://colab.ncsa.uiuc.edu/EESHG1/Documents/repository/conference/CyberinfrastructureTechQAQC.pdf)
 447 [chQAQC.pdf](http://colab.ncsa.uiuc.edu/EESHG1/Documents/repository/conference/CyberinfrastructureTechQAQC.pdf))
 448
- 449 Maidment, D. R. (Ed.), 2005. *Hydrologic Information System Status Report, Version 1*,
 450 Consortium of Universities for the Advancement of Hydrologic Science Inc, Washington, D.
 451 C., 224 p. (Available at <http://www.cuahsi.org/docs/HISStatusSept15.pdf>)
 452
- 453 Maidment, D. R. (Ed.), 2008. *CUAHSI Hydrologic Information System: Overview of Version*
 454 *1.1*, Consortium of Universities for the Advancement of Hydrologic Science Inc,
 455 Washington, D. C., 92 p. (Available at <http://his.cuahsi.org/documents/HISOverview.pdf>)
 456
- 457 Michener, W. K., Brunt, J. W., Helly, J. J., Kirchner, T. B., Stafford, S. G., 1997. Nongeospatial
 458 metadata for the ecological sciences, *Ecological Applications*, 7 (1), 330-342,
 459 doi:10.1890/1051-0761(1997)007[0330:NMFTEs]2.0.CO;2.
 460
- 461 Montgomery, J. L., Harmon, T., Kaiser, W., Sanderson, A., Haas, C. N., Hooper, R., Minsker,
 462 B., Schnoor, J., Clesceri, N. L., Graham, W., Brezonik, P., 2007. The WATERS Network:
 463 an integrated environmental observatory network for water research, *Environmental Science*
 464 *and Technology*, 41 (19), 6642-6647. (Available at
 465 http://pubs.acs.org/subscribe/journals/esthag/41/i19/pdf/100107feature_waters.pdf)
 466
- 467 Mourad, M., Bertrand-Krajewski, J. L., 2002. A method for automatic validation of long time
 468 series of data in urban hydrology, *Water Science & Technology*, 45 (4-5), 263-270.
 469
- 470 National Research Council, 2008. *Integrating Multiscale Observations of U.S. Waters*,
 471 Committee on Integrated Observations for Hydrologic and Related Sciences, Water Science
 472 and Technology Board, Division on Earth and Life Studies, The National Academies Press,
 473 Washington, D. C.
 474
- 475 National Science Foundation, 2007. *Cyberinfrastructure vision for 21st century discovery*,
 476 National Science Foundation Cyberinfrastructure Council, NSF 07-28. (Available at
 477 <http://www.nsf.gov/pubs/2007/nsf0728/index.jsp>)
 478

- 479 Sheth, A. P., Larson, J. A., 1990. Federated database systems for managing distributed,
480 heterogeneous, and autonomous databases, *Computing Surveys*, 22 (3),
481 doi:10.1145/96602.96604.
482
- 483 Valentine, D., Zaslavsky, I., Whitenack, T., Maidment, D. R., 2007. Design and implementation
484 of CUAHSI WATERML and WaterOneFlow web services, Extended Abstract, Published
485 Collection: Geoinformatics 2007 Bibliography. (Available at
486 http://gsa.confex.com/gsa/2007GE/finalprogram/abstract_122329.htm)
487
- 488 WATERS Network, 2008. Draft science, education, and design strategy for the WATER and
489 Environmental Research Systems Network, WATERS Network Project Office. (Available at
490 <http://www.watersnet.org/docs/SEDS-20080227-draft.pdf>)
491
- 492 Youn, C., Baru, C., Bhatia, K., Chandra, S., Lin, K., Memon, A., Memon, G., Seber, D., 2007.
493 GEONGrid portal: design and implementations, *Concurrency and Computation: Practice
494 and Experience*, 19 (12), 1597-1607, doi:10.1002/cpe.1129.
495
- 496 Zaslavsky, I., Valentine, D., Whiteaker, T., (Eds.), 2007. CUAHSI WaterML, OGC Discussion
497 Paper, OGC 07-041r1, Version 0.3.0. (Available at
498 <http://www.opengeospatial.org/standards/dp>)
499

500 **Figure Captions**

501 Figure 1. Little Bear River test bed monitoring site locations.

502

503 Figure 2. Little Bear River sensor network map.

504

505 Figure 3. ODM Streaming Data Loader wizard-based graphical user interface. The top panel
506 shows the listing of datalogger files that have been mapped and scheduled to be loaded into the
507 LBRTB ODM database. The bottom panel shows the interface for mapping the individual
508 columns in a single datalogger file to the ODM schema.

509

510 Figure 4. ODM Tools data editing interface.

511

512 Figure 5. Example specific conductance data series from the Paradise monitoring site before and
513 after quality control editing using ODM Tools.

514

515 Figure 6. Data discovery, visualization, and analysis components of the LBRTB observatory
516 information system.

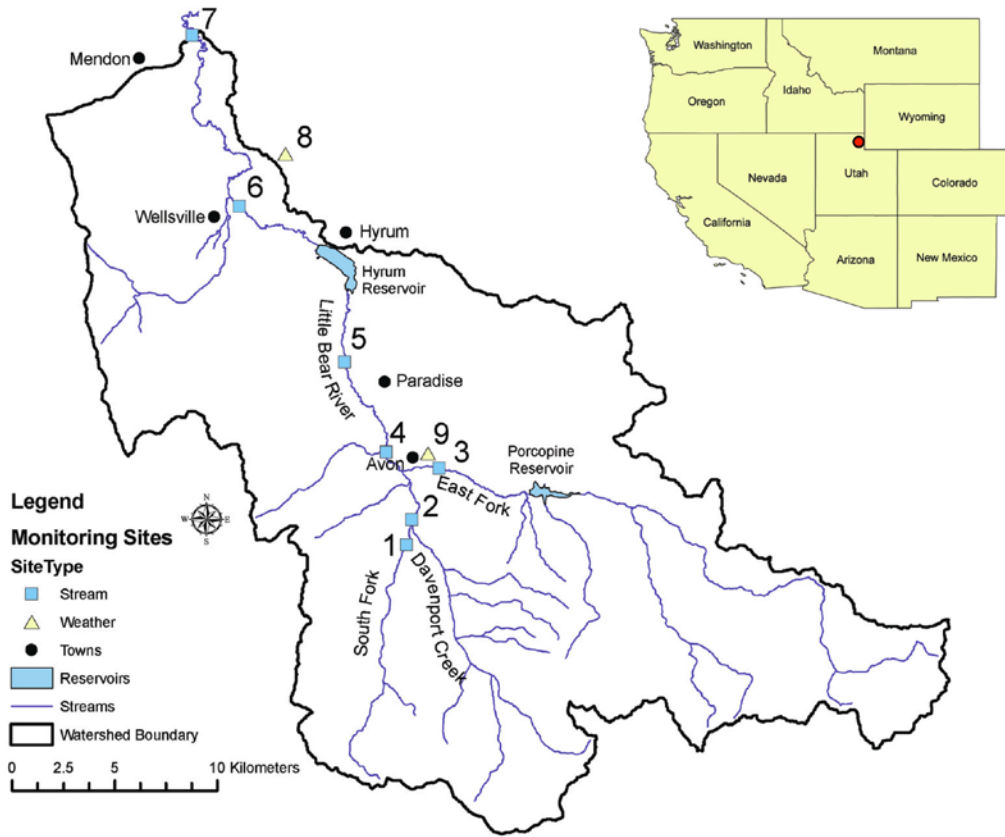
517 Table 1. Monitoring Sites in the LBRTB.

Site Number	Site Name	Latitude	Longitude	Site Type
1	Upper South Fork Little Bear River	41.4954	-111.818	Stream
2	Lower South Fork Little Bear River	41.5065	-111.8151	Stream
3	East Fork Little Bear River	41.5292	-111.7993	Stream
4	Little Bear River below Confluence of East and South Forks	41.5361	-111.8305	Stream
5	Little Bear River near Paradise	41.5756	-111.8552	Stream
6	Little Bear River near Wellsville	41.6435	-111.9176	Stream
7	Little Bear River near Mendon	41.7185	-111.9464	Stream
8	Lower Watershed Weather Station	41.667	-111.8906	Weather
9	Upper Watershed Weather Station	41.5355	-111.8059	Weather

518

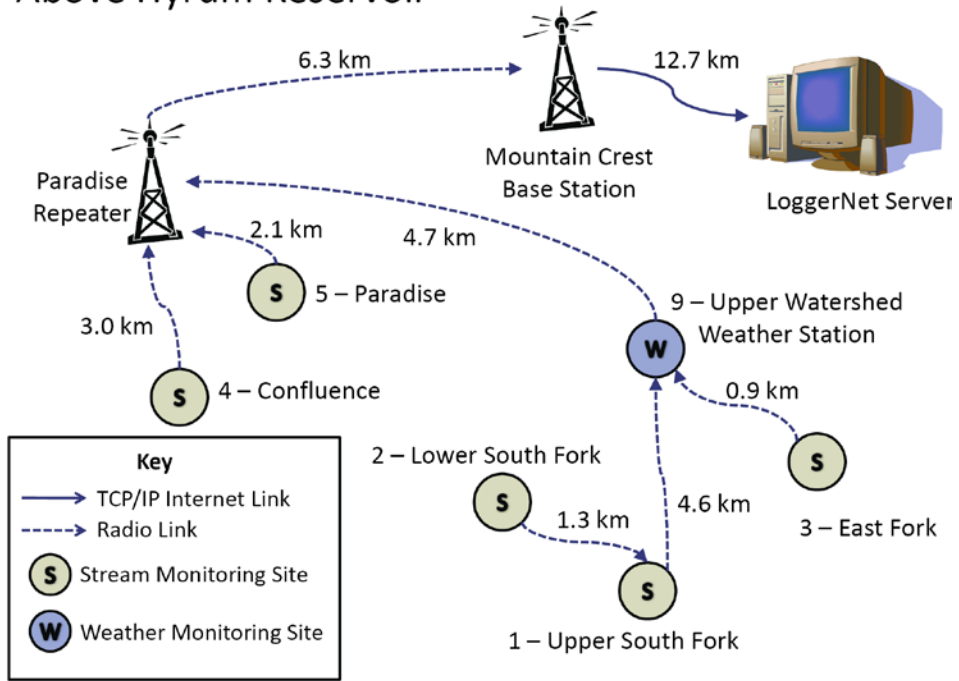
519 Table 2. Sensor Specifications for the LBRTB Monitoring Sites.

Site Type	Data Collection Frequency	Variable	Sensor
Stream	30 minutes	Water Temperature	Hydrolab MiniSonde 5 thermistor
		Dissolved Oxygen	Hydrolab MiniSonde 5 optical dissolved oxygen sensor
		pH	Hydrolab MiniSonde 5 reference electrode
		Specific Conductance	Hydrolab MiniSonde 5 four electrode conductivity sensor
		Turbidity	Forest Technology Systems DTS-12 Turbidity Sensor
		Stage	KWK Technologies SPXD-600 Pressure Transducer
Weather	1 hour	Air Temperature	Campbell Scientific CS215 temperature and relative humidity sensor
		Relative Humidity	Campbell Scientific CS215 temperature and relative humidity sensor
		Solar Radiation	Apogee PYR-P silicon pyranometer
		Precipitation	Texas Electronics TE25 tipping bucket
		Barometric Pressure	Setra 278 barometric pressure sensor
		Wind Speed	R. M. Young Wind Sentry Set
		Wind Direction	R. M. Young Wind Sentry Set



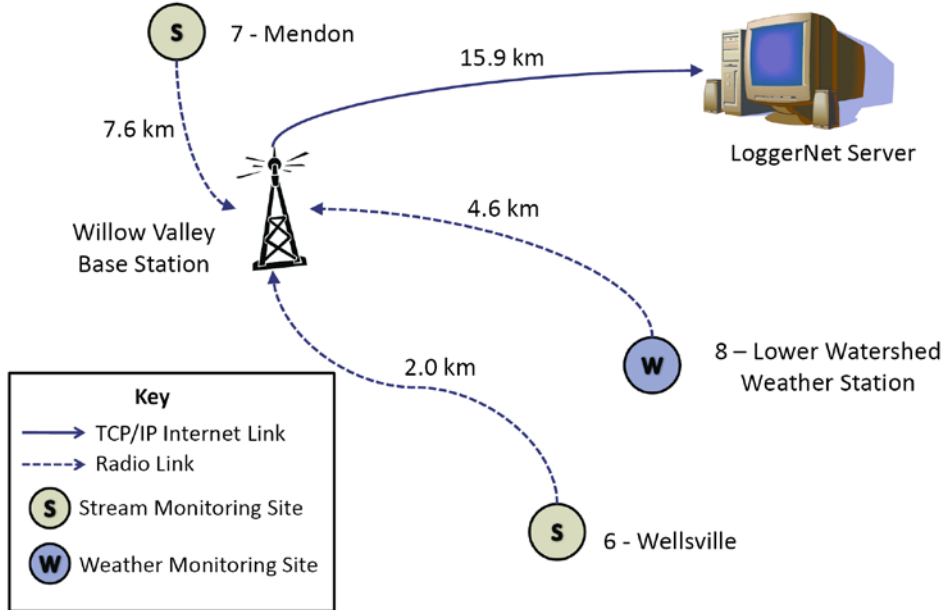
520
521 **Figure 1.** Little Bear River test bed monitoring site locations.

Above Hyrum Reservoir



522
523

Below Hyrum Reservoir



524
525

526 **Figure 2.** Little Bear River sensor network map.

ID	Server Address	Database Name	File Location Type	File Location	Schedule Period	Schedule Beginning	Last Update
11	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 days	8/13/2008 5:00...	9/28/2008 5:00...
13	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 days	8/13/2008 5:00...	9/28/2008 5:00...
14	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
15	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
2	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 days	8/13/2008 5:00...	9/28/2008 5:00...
7	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
8	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 days	8/13/2008 5:00...	9/28/2008 5:00...
9	(local)	LittleBearRiver	Local	C:\working\proje...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
1	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
16	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
17	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
5	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
10	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
6	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
12	(local)	LittleBearRiver	Local	\\brahma.uwrl.us...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...
18	(local)	LittleBearRiver	Web	http://waterdata...	1 hours	8/13/2008 5:00...	9/28/2008 5:00...

(a)

TIMESTAMP	RECORD	Batt_Volt_Min	Level_ft_Avg	Temp_degC_Avg	DO_mgL_Avg	DO_Perc_Avg	SpCond_mS_Avg	pH
2007-12-10 14:3...	1920	13.61078	0.7526631	8.19	10.29	104.4667	408.3333	8.09
2007-12-10 15:0...	1921	13.3236	0.7517405	8.188	10.174	103.3	408.4	8.10
2007-12-10 15:3...	1922	13.08524	0.750318	8.158333	10.06	102.1	408.6667	8.10
2007-12-10 16:0...	1923	12.94825	0.7486265	8.089999	9.94	100.65	409	8.08
2007-12-10 16:3...	1924	12.80958	0.7494722	8.003333	9.766666	98.71667	409.3333	8.06
2007-12-10 17:0...	1925	12.77627	0.748588	7.908333	9.608333	96.88333	410	8.04
2007-12-10 17:3...	1926	12.74848	0.7465889	7.808333	9.466666	95.21667	410.8333	8.00
2007-12-10 18:0...	1927	12.70958	0.7485111	7.703333	9.388333	94.19999	411.5	7.97
2007-12-10 18:3...	1928	12.69846	0.7479729	7.6	9.366666	93.75	412	7.95
2007-12-10 19:0...	1929	12.68743	0.7473962	7.485	9.373333	93.55	412	7.94
2007-12-10 19:3...	1930	12.64316	0.7479344	7.388333	9.381667	93.39999	412.5	7.93
2007-12-10 20:0...	1931	12.69291	0.7483574	7.275	9.428333	93.63333	412.1667	7.91
2007-12-10 20:3...	1932	12.62892	0.744513	7.216667	9.431666	93.48333	412.6667	7.91
2007-12-10 21:0...	1933	12.63609	0.7447436	7.16	9.448333	93.51666	412.6667	7.90

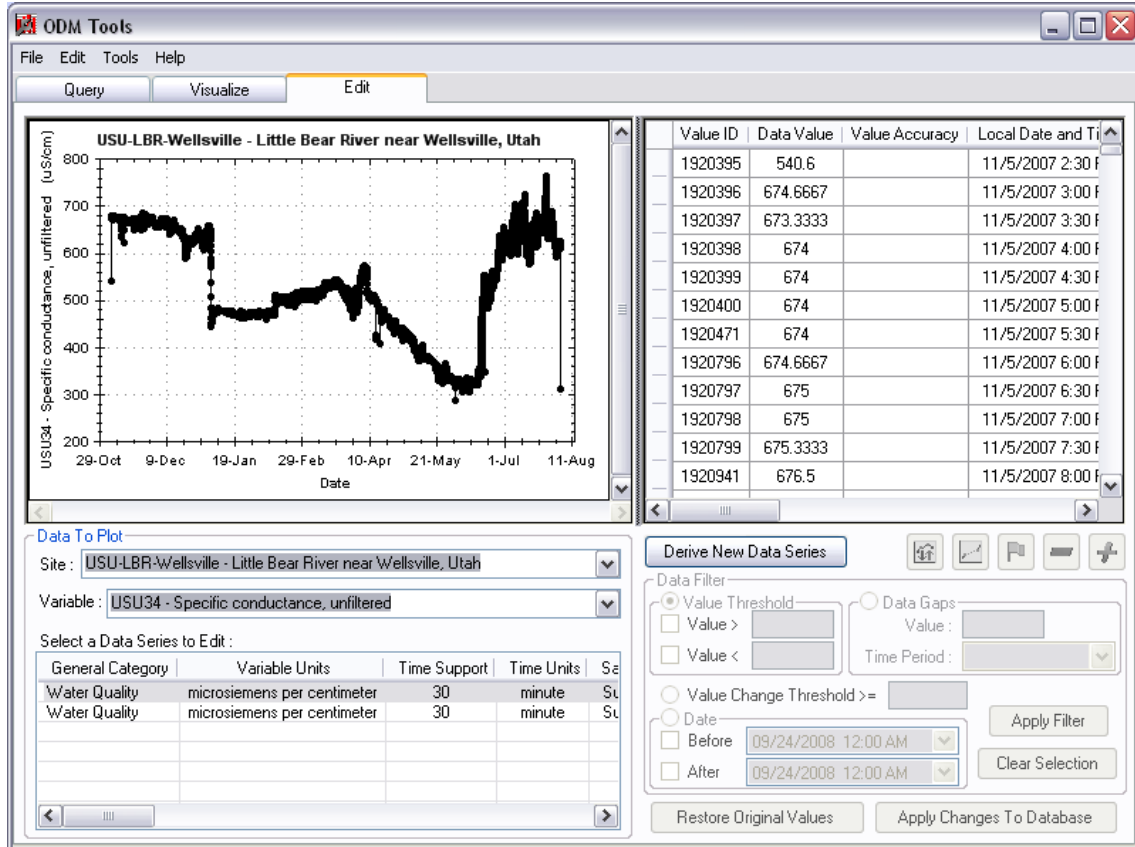
Value Column	Site	Variable	Offset Type	Offset Value	Method	Source
1	Batt_Volt_Min	7	3	<None>	<None>	22
2	DO_mgL_Avg	7	32	<None>	<None>	19
3	DO_Perc_Avg	7	33	<None>	<None>	19

(b)

527
528
529

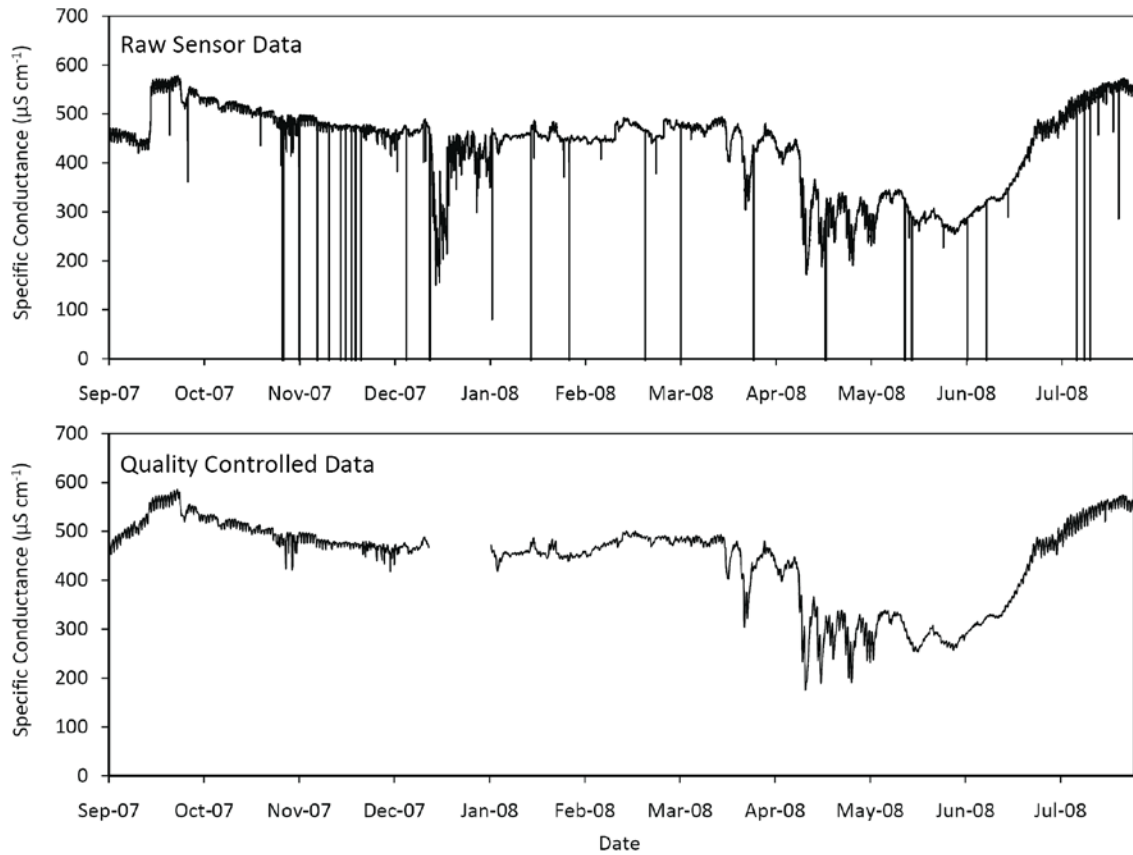
530
531
532
533
534
535
536

Figure 3. The ODM Streaming Data Loader wizard-based graphical user interface. Panel (a) shows the listing of datalogger files that have been mapped and scheduled to be loaded into the LBRTB ODM database. Panel (b) shows the interface for mapping the individual columns in a single datalogger file to the ODM schema.

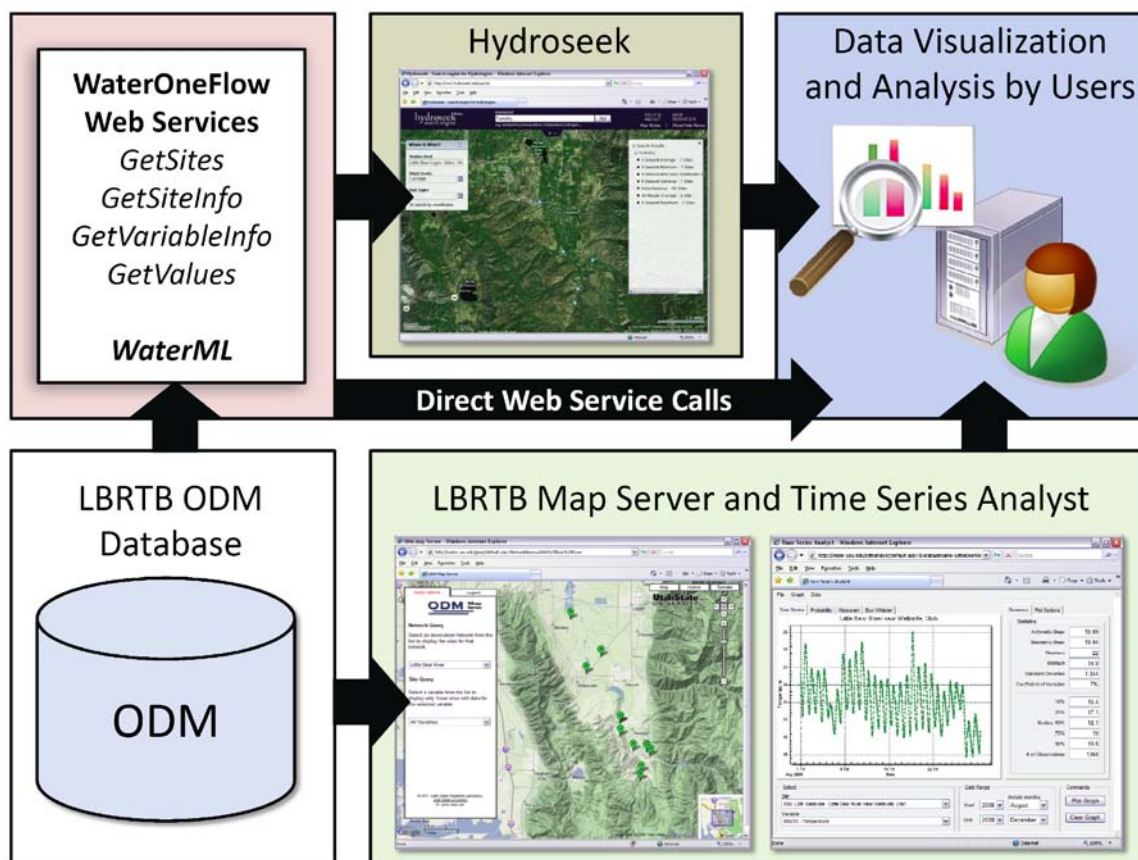


537
538
539

Figure 4. ODM Tools data editing interface.



540
541 **Figure 5.** Example specific conductance data series from the Paradise monitoring site
542 before and after quality control editing using ODM Tools.



543
544
545
546

Figure 6. Data discovery, visualization, and analysis components of the LBRTB observatory information system.