Extending HydroShare to enable hydrologic time series data as social media

Jeffrey M. Sadler, Daniel P. Ames and Shaun J. Livingston

ABSTRACT

The Consortium of Universities for the Advancement of Hydrologic Science Inc. (CUAHSI) hydrologic information system (HIS) is a widely used service oriented system for time series data management. While this system is intended to empower the hydrologic sciences community with better data storage and distribution, it lacks support for the kind of 'Web 2.0' collaboration and social-networking capabilities being used in other fields. This paper presents the design, development, and testing of a software extension of CUAHSI's newest product, HydroShare. The extension integrates the existing CUAHSI HIS into HydroShare's social hydrology architecture. With this extension, HydroShare provides integrated HIS time series with efficient archiving, discovery, and retrieval of the data, extensive creator and science metadata, scientific discussion and collaboration around the data and other basic social media features. HydroShare provides functionality for online social interaction and collaboration while the existing HIS provides the distributed data management and web services framework. The extension is expected to enable scientists to access and share both national- and laboratory-scale hydrologic time series datasets in a standards-based web services architecture combined with social media functionality developed specifically for the hydrologic sciences.

Key words | digital objects, metadata, social networks, time series data, web systems

Jeffrey M. Sadler (corresponding author)
Daniel P. Ames
Civil and Environmental Engineering,
Brigham Young University,
368 Clyde Building,
Provo,
UT 84602,
USA
E-mail: *jeffrey.sadler2@gmail.com*

Shaun J. Livingston
Applied Physics,
Brigham Young University,
368 Clyde Building,
Provo,
UT 84602,
USA

INTRODUCTION

Scientific data exchange

The emergence of the so-called fourth scientific research paradigm, data-intensive science, has brought with it significant challenges and opportunities that are magnified by rapid increases in the availability of data and data management tools (Piasecki *et al.* 2010; Tolle *et al.* 2011). The aggregation and distribution of heterogeneous data across disciplines enables scientists to address ever more complex challenges such as global climate variability, the effects of urbanization, and decreasing availability of water (Michener *et al.* 2012). Recent advances in water data cyberinfrastructure have drastically simplified the storage and exchange of water related data by individual researchers and large agencies alike (Botts *et al.* 2008; Beran & Piasecki 2009; Horsburgh *et al.* 2009; Morgenschweis 2011; Valentine *et al.* 2012; Conner *et al.* 2013; Tarboton *et al.* 2013). DataONE is an example of these systems as a platform designed to facilitate scientific discovery and sharing by cataloging existing data repositories (Michener *et al.* 2012).

Another system gaining in use is the Consortium of Universities for the Advancement of Hydrologic Sciences Inc.'s (CUAHSI) hydrologic information system (HIS). HIS is a service-oriented system designed for the storage and exchange of hydrologic time series data consisting of three parts (Tarboton *et al.* 2009; Ames *et al.* 2010, 2012). First, HydroServer hosts the data and provides a front-end interface to the database where the data are stored (Conner *et al.* 2013). HydroServer has been developed in two major versions: one based on the .NET Framework and Microsoft

doi: 10.2166/hydro.2015.331

SQL Server, and the other based on PHP and MySQL. The PHP/MySQL version of HydroServer accommodates the insertion of automatically transmitted sensor data via a web application programming interface (API) (Kadlec *et al.* 2015; Sadler *et al.* 2016) and supports OGC Web Feature Services (Michaelis & Ames 2012) and WaterML2 (Taylor *et al.* 2010). Second, HIS Central catalogs the time series data stored on registered HydroServers making them discoverable to client tools. And third, with HydroDesktop, the main client tool of HIS, users can discover, visualize, and analyze HIS time series data (Ames *et al.* 2012). To exchange the data between the components CUAHSI developed WaterOneFlow web services and the WaterML1.1 format.

Data sharing

CUAHSI HIS is an example data exchange service intended specifically for scientific data. While systems and technologies for scientific data exchange have been well-developed, major scientific advances may require more than the mere exchange of data. We suggest that a more complete and useful notion of scientific data sharing should include at least the following components:

1. A mechanism to efficiently archive, discover, and retrieve data (Piasecki *et al.* 2010).
2. Extensive creator and science metadata including versioning, licensing, and provenance information (Bechhofer *et al.* 2013).
3. A persistent digital identifier such as a digital object identifier (DOI) (Paskin 2008).
4. A system for facilitating scientific discussion and collaboration around the data.

In the hydrologic sciences, much has already been accomplished regarding the first three components of this list, as evidenced by the volume of research already cited herein. However, little headway has been made in terms of facilitating the open scientific discussion and collaboration around data. One notable exception is the recent policy of the online journal, PLOS ONE, requiring all authors to make their data publicly available upon publication (PLOS ONE 2014). This data publishing requirement combined with the open discussion model supported by the journal web site (readers can post comments on any article) provides a framework for open collaboration on data associated with publications.

A similar collaboration and data sharing support system for data associated with pre-publication research has the potential of increasing the speed of scientific advances and the related publication process (Tarboton *et al.* 2013). An example of such a system is iPlant in which users can discuss any file or folder in a structured comment management system (Goff *et al.* 2011). Similarly, D4Science.org users can share 'updates' to which datasets can be attached on a public news feed. These updates can be discussed and 'favorited' (Assante *et al.* 2014). These systems provide scientists with a simple workspace for discussion and clarification regarding individual or groups of datasets which are not necessarily associated with published journal articles.

HydroShare as a social media platform

The HydroShare project is designed to provide hydrologic scientists each of the four functionalities of a data-sharing network described above, and notably emphasizes the last point, providing a space for collaboration and discussion. HydroShare, an NSF-funded partner project to CUAHSI HIS, is a data sharing system for static datasets (e.g. files on disk rather than dynamic, database-driven HIS datasets) built on the Django web application framework (www.djangoproject.com/) and the Integrated Rule-Oriented Data System (iRODS) (http://irods.org/). It is focused on providing a simple web-based data sharing platform for hydrologists to discover and access hydrologic data and models (Tarboton *et al.* 2014).

The design of HydroShare is based around the concept of a 'resource' as a basic unit of data. A resource may be a single observation (e.g. an aerial image or single streamflow measurement) or a collection of data (e.g. a NetCDF file with multiple observations or a streamflow time series). Each resource has an associated resource type, a number of which have been proposed and/or developed for HydroShare including generic resource, geographic feature resource, static time series resource, raster data resource and aggregate resource. Resource types are contained in Django apps making the Django framework very suitable for HydroShare's extensible design. By supporting specific resource types, HydroShare can provide relevant value added functionality and metadata support (e.g. an online map viewer and projection metadata for a raster data resource). HydroShare uses the Dublin Core metadata standard (www.dublincore.org/documents/dcmi-terms) as the base metadata definition for all resource types. HydroShare is similar to other file sharing systems (e.g. FTP, DropBox), although highly augmented with structured metadata, data type specific value added functionality, and social networking capabilities.

HydroShare's social media functionality provides a system for scientific discussion and collaboration. In HydroShare, participating scientists create user profiles which can include a profile picture, a listing of research interests, and attached CV. A HydroShare user has several access control options for his or her resources. HydroShare also provides a structured commenting environment with the option of endorsing or 'liking' comments. A HydroShare user can comment on all resources to which he or she has access. When a user comments on a resource, the owner of that resource will receive an email notification. User profiles, groups, and informal communication around HydroShare resources facilitate an open environment of scientific discussion and collaboration (Tarboton *et al.* 2014).

Kaplan & Haenlein (2010) characterize social media according to level of self-presentation/self-disclosure and social presence/media richness in terms of six categories: blogs, social networks, virtual social worlds, collaborative projects, content communities, and virtual game worlds. HydroShare's social media design combines elements of social networks (creation of groups, and a user profile which allows for professional self-disclosure), collaborative projects (collaboration around content to produce a more complete and better result), and content communities (upload, sharing, and discoverability of research data). This combination of social network, collaborative project, content community is uniquely suited to aid hydrologists in their individual and collaborative research.

Melding HIS and HydroShare

While the current HIS presents a useful solution to the challenge of storing and exchanging time series data it lacks the support for web-based interaction with these datasets in a system that supports social media-style scientific collaboration and discussion. The remainder of this paper presents the design, development, and testing of a new software extension intended to fill the need for web-based HIS functionality, online collaboration, and social networking for HIS time series data (turning a HIS time series dataset into social media). This is accomplished by extending the HydroShare system by creating a new resource type unique from any of the existing or proposed HydroShare resource types: a HIS Referenced Time Series resource type – a resource that, rather than storing static data files on disk, retrieves data dynamically from one of the many existing HydroServers in the HIS network (Figure 1). As HydroShare resources, HIS time series will be digital social objects that can be discussed, endorsed, followed, and shared – effectively becoming social media. Furthermore, the referenced HIS time series data will be provided the HydroShare data curation features such as provenance and attribution metadata.

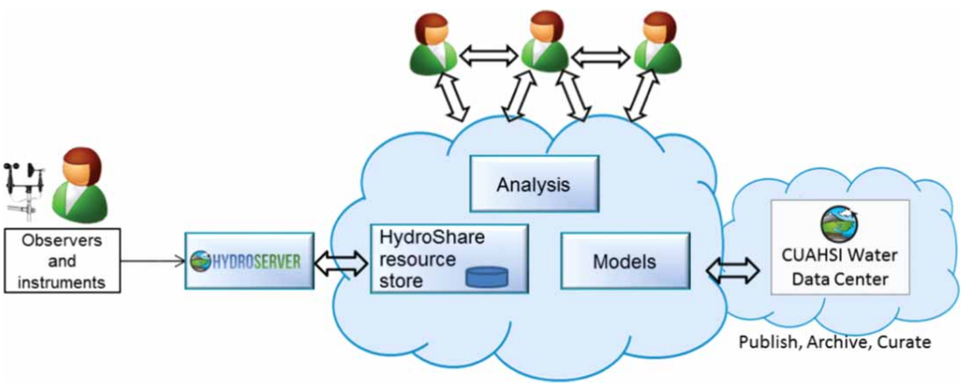

Figure 1 | Overall workflow of a HIS Referenced Time Series resource (Tarboton *et al.* 2014).

The testing of this software extension was done in reference to stream stage data collected on the Provo River in Provo Utah, USA. The data were collected using low-cost data logging and transmission hardware, stored on a Hydroserver (Sadler *et al.* 2016) and added to HydroShare using our software extension. Through the social media functionality provided by HydroShare, the owner of the data is able to receive input from select colleagues to determine whether the data are reasonable and can be relied upon for hydrologic model calibration or validation.

METHODS

Design of a Referenced Time Series data type for HydroShare

This section describes the design of the HIS Referenced Time Series resource type. The differences between the HIS Referenced Time Series resource data type and generic resource data type consist primarily of differences in metadata and resource creation. The metadata differences are reflected in the HIS Referenced Time Series data model. The HIS Referenced Time Series data model consists of three main categories of classes: the class for the resource data type (RefTimeSeries), the class for the resource data type extended metadata (RefTSMetadata), and the classes for the individual extended metadata terms (Method, QualityControlLevel, Variable, Site). The class structure for the HIS Referenced Time Series data type is depicted in Figure 2.

The RefTimeSeries class inherits from the HydroShare AbstractResource class. This way HIS Referenced Time Series objects will have universally needed attributes such as 'short_id' and 'doi.' The RefTSMetadata class inherits from the HydroShare CoreMetaData class and is the link between the RefTimeSeries class and the metadata term classes. In RefTSMetadata class 'One-To-One' relations are made with each of the extended metadata classes (ReferenceURL, Method, QualityControlLevel, Variable, and Site). The individual metadata term classes are then included as supported metadata elements for the HIS Referenced Time Series resources. This way the terms' class methods can be used to create, update, and delete class instances associated with HIS Referenced Time Series resources. A key method of the RefTSMetadata and CoreMetaData classes is 'get_xml.' This method writes all of the metadata terms and their values into a science metadata XML document. The CoreMetadata 'get_xml' method does this for the Dublin Core standard metadata elements while the RefTSMetadata 'get_xml' method must be written to do this for each of the extended metadata terms for the class instance.

Finally a class is made for each of the extended metadata terms. These classes inherit from the HydroShare AbstractMetaDataElement class. The AbstractMetaDataElement has the 'create,' 'update,' and 'remove' methods. These methods however simply redirect to the child class methods and thus must be implemented in the child class. In these classes there is one required attribute: 'term'. Other attributes needed for further description can be added. For example, the Site class has the 'term,' 'name,' 'code,' 'latitude,' and 'longitude' attributes. Example code for these classes has been included in the Appendix (available with the online version of this paper).

To populate the extended metadata terms, the HIS Referenced Time Series data type requires a unique resource creation process. In contrast to the creation of a generic HydroShare resource, no files are uploaded when creating a HIS Referenced Time Series resource. Rather, a reference is needed that links to data stored on an externally hosted HydroServer. Additionally, to support WaterOneFlow services, users need a way to select a specific site and relevant variable. These differences in creation are accommodated by the HydroShare generic resource creation workflow which can redirect to a different page depending on the selected data type. In the generic resource creation workflow the 'pre_create_resource' signal is sent when the creation process is first begun. This signal accepts an optional reference for a redirect page which the HIS Referenced Time Series 'pre_create_resource' receiver returns. The system then redirects the user to the page where he or she can create an HIS Referenced Time Series resource with interactive controls for selecting a site and variable.

In the HIS Referenced Time Series resource creation page, the user is provided a list of all HydroServer URLs which are cataloged in HIS Central. There is also the possibility of providing a reference URL for a HydroServer which

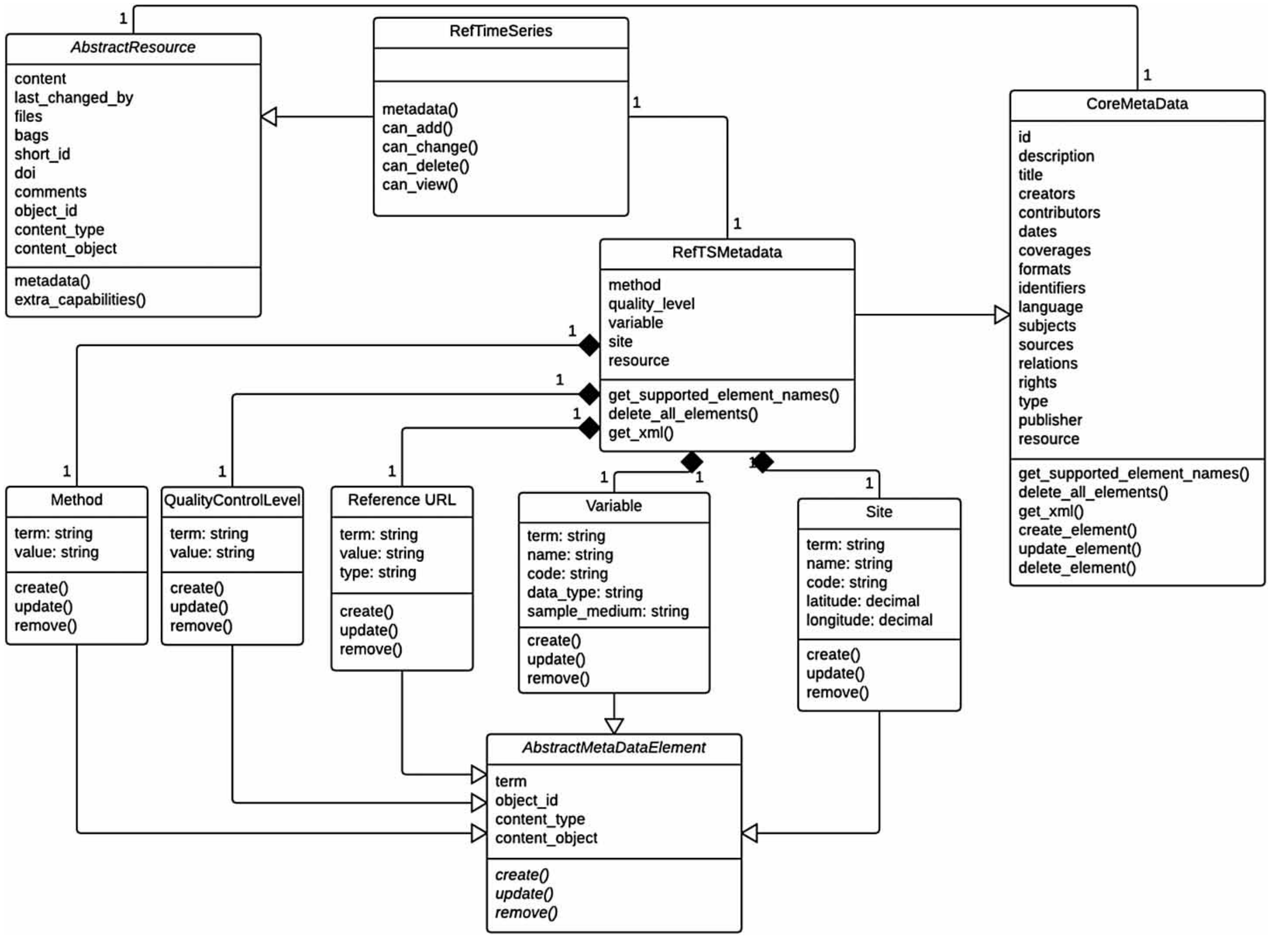

Figure 2 | UML class diagram for HIS referenced time series data type.

has not been cataloged with HIS Central. If the user supplies a SOAP URL, HydroShare uses the WaterOneFlow methods GetSites and GetSiteInfo to retrieve available sites, and then the available variables for the selected site. When the site and variable have been selected the GetDataValues method is executed and the time series data is retrieved. The user also has the option of providing a REST URL. Because the REST URL contains all information needed to access a single time series, the refinement needed with a SOAP URL is unnecessary. Once the time series has been successfully retrieved, the extension constructs and displays a preview graph. A screen shot of the creation page is shown in Figure 3.

Once the resource has been created, our extension provides a link to download the referenced time series in WaterML2 and CSV formats. If the original data are in WaterML1.1, this format is also available. These files are constructed and downloaded on the fly and are therefore always up to date with the data source. The landing page for a HIS Referenced Time series is shown in Figure 4.

Case study validation

The implementation of the design of the HIS Referenced Time Series within the HydroShare-Django framework is validated with respect to stream stage data collected on the Provo River shown in Figure 5. The data were stored on a HydroServer which is not registered with the HIS Catalog. The researcher is using the Lower Provo River watershed as a study area for urban hydrology and would like to eventually use data from this station to calibrate his model. Because the data were collected using low-cost data-logging and transmission hardware

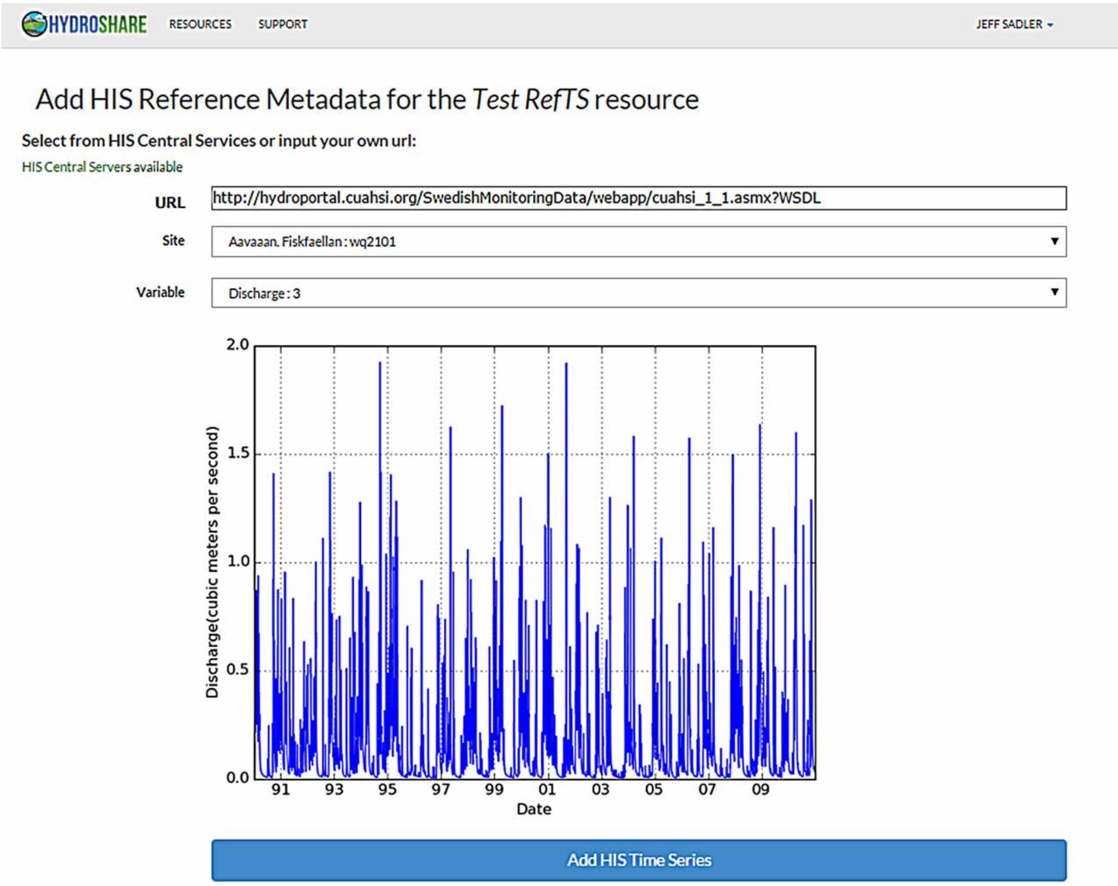

Figure 3 | HIS referenced time series creation page.

(Sadler *et al.* 2016), instead of vetted, commercial equipment, the researcher has some questions about the validity, and therefore utility, of the collected data. The data gaps seen in Figure 5 contribute to this questioning. The researcher is especially interested in the peak data points on 23 August 2014. Since these values are quite abnormal compared to the rest of the data points, the researcher would like some verification from certain colleagues that the values are reasonable and it was not simply an error in their collection or transmission. If the data are accurate, the recorded storm event can be used in his hydrologic model calibration. Since he is unsure about the data's accuracy, the researcher would like to share the data with only a select few colleagues who have their own resources which would be helpful in resolving his question. With the HydroShare extension we created, the researcher should be able to do the following: (1) create a HIS Referenced Time Series Resource by referencing the time series stored on a HydroServer which is not cataloged in HIS Central, (2) share the resource with only select colleagues, (3) comment on and receive colleague's comments on the created HIS Referenced Time Series resource, (4) endorse comments and see his or her colleague's endorsements of comments, and (5) finally, with the help of the colleagues' collaboration, make a decision whether or not to accept the collected data, make the resource public, and continue using this station for his modeling needs.

RESULTS AND DISCUSSION

This section presents the results and discussion of the implementation of the Referenced Time Series resource type in HydroShare. The results are based upon the HydroShare system found at www.hydroshare.org. Source code for

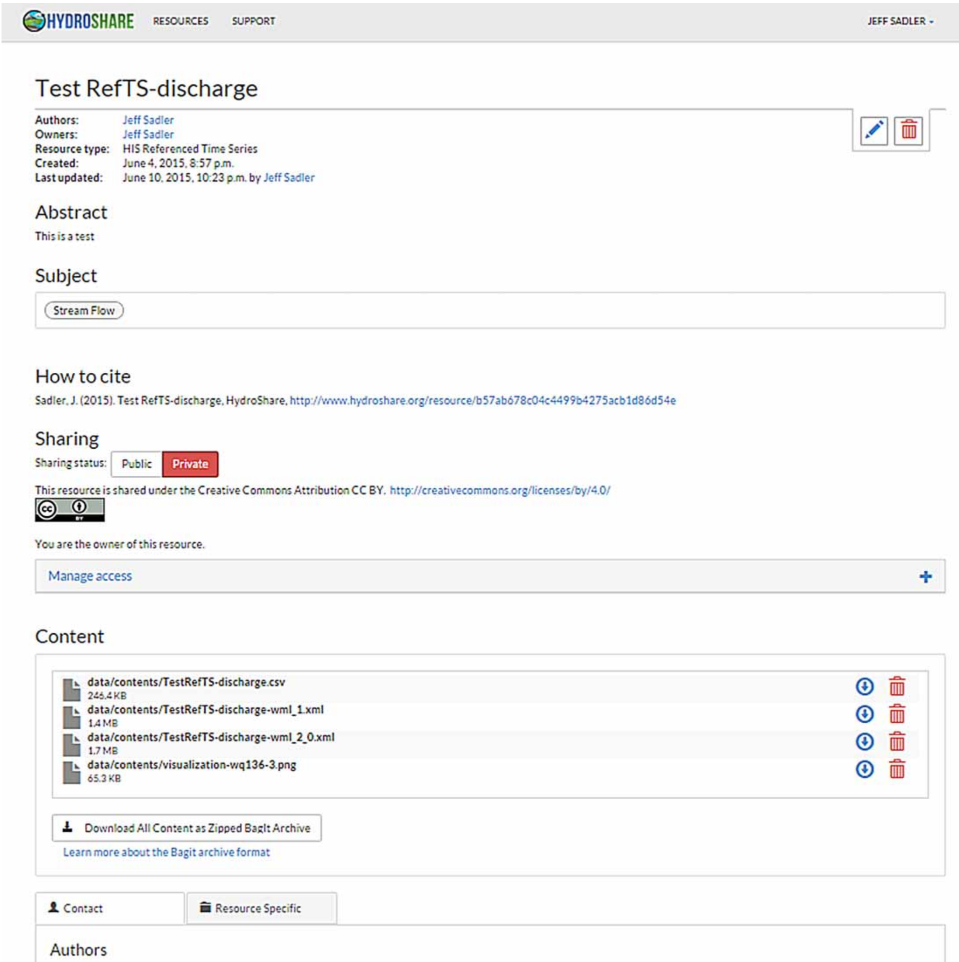

Figure 4 | Resource Landing Page for HIS Referenced Time Series.

the HIS Referenced Time Series resource is available together with full HydroShare source code at http://github.com/hydroshare/hydroshare.

Functional testing was undertaken to ensure that the software worked as designed and as described above under the section 'Case study validation'. With the help of the comments of the researcher's colleagues, shown in Figure 6, he was able to feel confident about the accuracy of the collected data. The successful implementation of the HIS Referenced Time Series in HydroShare provides several equivalent or enhanced functionalities as compared to the original HIS and other generic and scientific data sharing platforms. These are summarized in Table 1. First, each HIS Referenced Time Series in HydroShare is available in the fully internationalized standard WaterML2 format while many current HIS time series are available only in the WaterML1.1 format. Each of the other systems only offers the data in the uploaded file format.

Second, like most of the systems we reviewed, there is much more flexibility regarding with whom and what portion of data is shared when using the HIS Referenced Time Series HydroShare resource as compared to the current HIS. In the current HIS only datasets from HydroServers registered with HIS Central are publicly discoverable and all of the datasets on a HydroServer are completely public or completely private. In contrast, a HIS Referenced Time Series is an individual dataset and can be shared with a select person or persons. Furthermore, a HydroShare user can create and share a HIS Referenced Time Series dataset from either public, registered HydroServers or unregistered HydroServers. This raises a

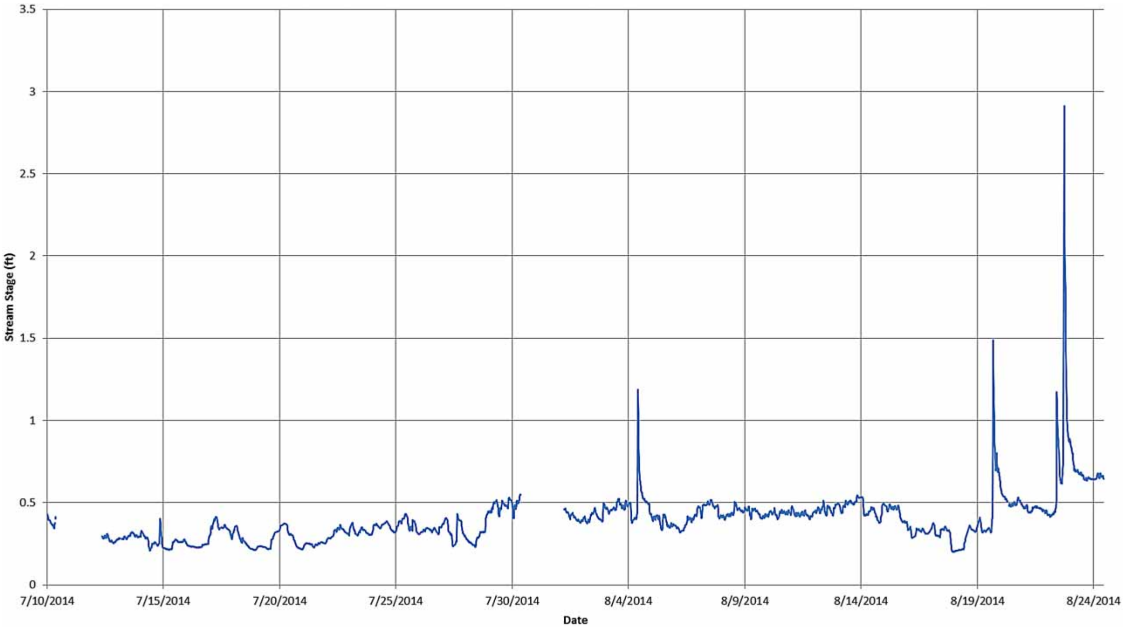

Figure 5 | Provo River stream stage 7/10–8/24/2014.

possible scientific attribution issue because a HydroShare user who did not collect and/or is not responsible for a dataset could publish and share it on the HydroShare system. However, the attribution metadata associated with the original dataset stored on HydroServer are maintained in read-only format in HydroShare. This may become a problem that needs to be monitored as usage grows. HydroServers requiring login credentials cannot be accessed to create a HydroShare HIS Referenced Time Series.

Third we considered the modes of discovery offered by each system. In the current HIS, the main discovery tool is HydroDesktop with which users can discover datasets according to geographic location, variable, and time frame (Ames *et al.* 2012). Similar functionality is provided by the DataONE platform. In HydroShare, on the other hand, users provide their own key words to associate with the resource. This is similar to PLOS ONE, D4Science.org and FigShare. Currently this is the most viable way of making HydroShare resources discoverable as queries with respect to location, variable, and time, like those in HydroDesktop, have not been implemented.

As a HydroShare resource the HIS time series can be commented on. Some motivations for users to comment on a HIS Referenced Time Series resource may include asking the resource owner clarification questions about the source or quality of the data, asking about any related data, or offering additional relevant information to the resource owner/research community. Before it is made public only those HydroShare users to whom the resource owner has given access can comment on the dataset. When public everyone, even anonymous users, will be able to add comments. This functionality is intended to promote a free and open collaborative environment. As such, there is currently no restriction on a user posting off-topic and/or inflammatory comments. While it is difficult to know the extent to which this type of 'trolling' activity will occur, the ability to endorse comments may act as a natural filter. If this is not sufficient, CUAHSI may need to provide a more direct solution. Our system is the only one that we reviewed that provides the ability to comment on both private and public resources.

As HIS Referenced Time Series resources, HIS datasets can benefit from third-party applications accommodated by HydroShare through a web API. Similar APIs are provided for most of the systems we reviewed. The HydroShare web API provides for web-based applications a similar resource that the HydroDesktop plug-in architecture provided for desktop developers (Ames *et al.* 2012). An interactive WaterML2 time series viewer has already been developed

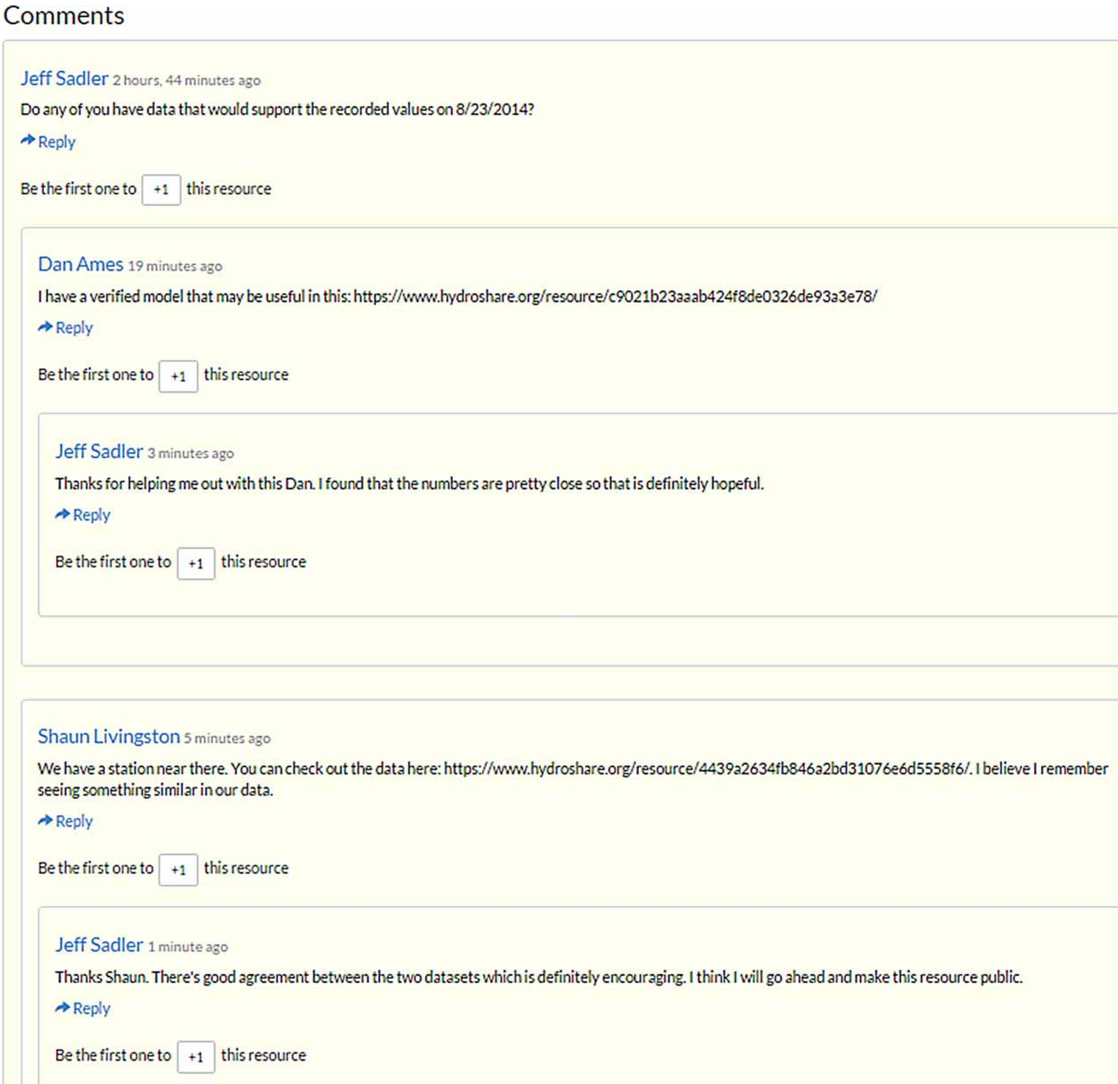

Figure 6 | Comments from use case researcher and colleagues.

using the Tethys platform (Swain *et al.* 2014) that, when deployed, will be able to be launched from the HIS Referenced Time Series resource landing page. This is only the first of many such third-party tools which will add even more functionality and value to the HIS Referenced Time Series resources.

The reference-based architecture of our software extension is unique across all of the systems we reviewed. Because the HIS Referenced Time Series resource data type was designed to simply reference existing data repositories, no data or HIS Referenced Times Series specific metadata (site, variable, etc.) editing is made available through HydroShare. The HydroShare resource simply reflects what is found in the HydroServer. Users can however, add more descriptive metadata with respect to the Dublin Core terms which are standard across all HydroShare resource data types. As a reference data type, the actual data can change. These changes are reflected in the HydroShare resource but prevent the assignment of a permanent digital identifier (DOI) to the actual data values. However, a DOI can be given to a HIS Referenced Time Series resource which would then refer to the data source: the reference URL that houses the data. It has been proposed to include the option of converting a HIS

Table 1 | Comparison between Existing HIS and the HydroShare HIS Referenced Time Series Data Type

	HydroShare Ref. Time Series (this paper)	HIS HydroServer Time Series (Tarboton et al. 2009)	DropBox (Drago et al. 2012)	PLOS ONE (MacCallum 2006)	iPlant (Goff et al. 2011)	D4Science.org (Assante et al. 2014)	FigShare (Singh 2011)	DataONE (Michener et al. 2012)
Retrieval format	WML2, CSV and some WML1.1	Some WML1.1 & some WML2	Uploaded file format	Uploaded file format	Uploaded file format	Uploaded file format	Uploaded file format	Native file format
Selective sharing	Yes	No	Yes	No	Yes	Yes	Yes	No
Public resource collection	Yes	Yes	No	Yes	No	Yes	Yes	Yes
Discoverability	User-defined key words	Spatial location, variable, and time frame	N/A	Machine Aided Indexing identified key words	N/A	User-defined key words	User-defined key words	Geographic bounding box, date range, key words
Commenting: public resources	Yes	No	N/A	Yes	N/A	Yes	Yes	No
private resources	Yes	No	Yes	N/A	No	No	No	No
Third-party application support	Web API	Desktop-based	Web API	Web API (retrieve only)	Web API	None	Web API	Web API
Automatic metadata extraction	Yes	N/A	No	No	No	No	No	No
Standard time series specific metadata	Yes	Yes	No	No	No	No	No	Yes

Referenced Time Series into a static time series which would take the latest data from the HydroServer and then eliminate the updating procedure; as a static time series, the resource would be able to be published with a DOI that refers to the data values themselves.

CONCLUSIONS

We successfully designed, integrated, and tested the HIS Referenced Time Series data type for the HydroShare system. With the HIS Referenced Time Series data type users can link time series data found on HydroServers into the HydroShare system. HydroShare social media functionality such as sharing, commenting, and endorsing HIS Referenced Time Series resources promote clarification and discussion which can enhance and accelerate collaborative science and modeling. HydroShare also provides HIS time series thorough science and creator metadata support.

In addition to the current HydroShare features which enhance their use and sharing, planned future developments will add even more value to HIS Referenced Time Series data. In future HydroShare versions, the ability to 'follow' a HIS Referenced Time Series resource as it changes will add value to the data. Additionally, as HydroShare begins to support the pre-processing, post-processing, and execution of hydrologic models, users will be able to easily integrate HIS Referenced Time Series resources as the data and workflows will all be contained conveniently in one system.

While the extension described in this paper is a fundamental step in integrating the existing HIS, it is unlikely that a scientist with many datasets would add each one individually. Scaling this model up may also clutter the HydroShare system. To make the integration of HIS time series more convenient, expanding the HIS Referenced Time Series model to include multiple datasets (i.e. more than one site or variable) or even an entire HydroShare has been proposed for future development. Similarly, further developments such as being able to subset datasets, annotate specific data points and view and/or publish as a new resource a comparison of multiple time series on the same plot may be added to the HIS Referenced Time Series resource. These enhancements could be developed in the HydroShare system itself or by third-party applications who would access HydroShare through the web API. Such developments would add considerably to the collaborative capabilities of the software.

Beyond CUAHSI HIS, there are many repositories that offer their time series data in a standardized way including research-, national-, and international-level repositories such as DataONE, the United States Geologic Survey, and the Global Earth Observation System of Systems (Butterfield *et al.* 2008). The HIS Referenced Time Series data type is expected to serve as a model which will be expanded and modified to support referencing time series data stored in these and other repositories. In this way the unique benefits that we have described for HIS time series datasets may be extended to many other time series datasets.

The connection between HydroShare and the CUAHSI HIS through the mechanism of the new HIS Referenced Time Series resource presented in this paper, provides hydrologists with a new, functional system for sharing and collaborating around HIS time series data – effectively turning HIS data into social media. This collaborative tool will help scientists leverage the ever-increasing amount of data to better understand the complex water challenges facing society today.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation under collaborative grants OCI-1148453 and OCI-1148090 for the development of HydroShare (www.hydroshare.org). 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., Kadlec, J. & Horsburgh, J. 2010 HydroDesktop: a free and open source platform for hydrologic data discovery, visualization, and analysis. In: *2010 AWRA Spring Specialty Conference: Geographic Information Systems (GIS) and Water Resources VI American Water Resources Association*, TPS-10-1, p. 1–882132.

Ames, D. P., Horsburgh, J. S., Cao, Y., Kadlec, J., Whiteaker, T. & Valentine, D. 2012 Hydrodesktop: web services-based software for hydrologic data discovery, download, visualization, and analysis. *Environ. Model. Softw.* **37**, 146–156.

Assante, M., Candela, L., Castelli, D., Mangiacrapa, F. & Pagano, P. 2014 A social networking research environment for scientific data sharing: the D4Science offering. *Grey J.* **10** (2), 65–71.

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., Cruickshank, D., Delderfield, M. & Dunlop, I. 2013 Why linked data is not enough for scientists. *Future Gener. Comput. Syst.* **29** (2), 599–611.

Beran, B. & Piasecki, M. 2009 Engineering new paths to water data. *Comput. Geosci.* **35** (4), 753–760.

Botts, M., Percivall, G., Reed, C. & Davidson, J. 2008 OGC® Sensor web enablement: overview and high level architecture. In: S. Nittel, A. Labrinidis & A. Stefanidis (eds). *Geosensor Networks*. Lecture Notes in Computer Science, Vol. 4540. Springer, Berlin Heidelberg, pp. 175–190.

Butterfield, M. L., Pearlman, J. S. & Vickroy, S. C. 2008 A system-of-systems engineering GEOSS: architectural approach. *Syst. J. IEEE* **2** (3), 321–332.

Conner, L. G., Ames, D. P. & Gill, R. A. 2013 Hydroserver Lite as an open source solution for archiving and sharing environmental data for independent university labs. *Ecol. Inform.* **18**, 171–177.

Drago, I., Mellia, M., Munafo, M., Sperotto, A., Sadre, R. & Pras, A. 2012 Inside dropbox. In: *Proceedings of the 2012 ACM conference on Internet measurement conference – IMC '12*. ACM Press, New York, USA, p. 481.

Goff, S. A., Vaughn, M., McKay, S., Lyons, E., Stapleton, A. E., Gessler, D., Matasci, N., Wang, L., Hanlon, M. & Lenards, A. 2011 The iPlant collaborative: cyberinfrastructure for plant biology. *Front. Plant Sci.* **2**, 34.

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. *Environ. Model. Softw.* **24** (8), 879–888.

Kadlec, J., StClair, B., Ames, D. P. & Gill, R. A. 2015 WaterML R package for managing ecological experiment data on a CUAHSI hydroServer. *Ecol. Inform.* **28**, 19–28.

Kaplan, A. M. & Haenlein, M. 2010 Users of the world, unite! the challenges and opportunities of social media. *Bus. Horiz.* **53** (1), 59–68.

MacCallum, C. J. 2006 ONE For all: the next step for PLoS. *PLoS Biology* **4** (11), e401.

Michaelis, C. D. & Ames, D. P. 2012 Considerations for implementing OGC WMS and WFS specifications in a desktop GIS. *J. Geog. Inform. Syst.* **4**, 161–167.

Michener, W. K., Allard, S., Budden, A., Cook, R. B., Douglass, K., Frame, M., Kelling, S., Koskela, R., Tenopir, C. & Vieglais, D. A. 2012 Participatory design of DataONE – enabling cyberinfrastructure for the biological and environmental sciences. *Ecol. Inform.* **11**, 5–15.

Morgenschweis, G. 2011 *Datenerfassung und fernübertragung Hydrometrie*. Springer, Berlin, pp. 513–535.

Paskin, N. 2008 Digital object identifier (DOI) system. *Encyclopedia of Library and Information Sciences* **3**, 1586–1592.

Piasecki, M., Ames, D., Goodall, J., Hooper, R., Horsburgh, J., Maidment, D., Tarboton, D. & Zaslavsky, I. 2010 Development of an information system for the hydrologic community. In: *Proceedings of the Ninth International Conference on Hydroinformatics*, Tianjin, China.

PLOS ONE 2014 Data availability. http://journals.plos.org/plosone/s/data-availability. Retrieved 11 October 2015.

Sadler, J. M., Ames, D. P. & Khattar, R. 2016 A recipe for standards-based data sharing using open source software and low-cost electronics. *J. Hydroinform.* **18** (2), 185–197.

Singh, J. 2011 Figshare. *J. Pharmacol. Pharmacother.* **2** (2), 138.

Swain, N., Christensen, S., Jones, N. & Nelson, E. 2014 Tethys: A Platform for Water Resources Modeling and Decision Support Apps. In: *AGU Fall Meeting Abstracts*, Vol. 1, p. 0760.

Tarboton, D. G., Horsburgh, J. S., Maidment, D. R., Whiteaker, T., Zaslavky, I., Piasecki, M., Goodall, J., Valentine, D. & Whitenack, T. 2009 Development of a Community Hydrologic Information. In: R. S. Anderssen, R. D. Braddock & L. T. H. Newham (eds). *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, pp. 988–994.

Tarboton, D., Idaszak, R., Horsburgh, J., Ames, D., Goodall, J., Band, L., Merwade, V., Couch, A., Arrigo, J. & Hooper, R. 2013 HydroShare: an online, collaborative environment for the sharing of hydrologic data and models. In: *AGU Fall Meeting Abstracts*, Vol. 1, p. 1510.

Tarboton, D., Idaszak, R., Horsburgh, J., Heard, J., Ames, D., Goodall, J., Band, L., Merwade, V., Couch, A. & Arrigo, J. 2014 HydroShare: Advancing Collaboration through Hydrologic Data and Model Sharing. In: D. P. Ames & N. Quinn (eds). *International Environmental Modelling and Software Society (iEMSs) 7th International Congress on Environmental Modelling and Software*, San Diego, California, USA. www.iemss.org/society/index/php/iemss-2014-proceedings.

Taylor, P., Walker, G., Valentine, D. & Cox, S. 2010 WaterML2. 0: harmonising standards for water observations data. In: *EGU General Assembly Conference Abstracts*, Vol. 12, p. 7680.

Tolle, K., Tansley, D. & Hey, A. 2011 The fourth paradigm: data-intensive scientific discovery. *Proc. IEEE* **99**, 8.

Valentine, D., Taylor, P. & Zaslavsky, I. 2012 WaterML, an information standard for the exchange of in-situ hydrological observations. In: *EGU General Assembly Conference Abstracts*, Vol. 14, p. 13275.

Winter, L. 2014 PLOS ONE to require rublic access for data. IFL Science http://www.iflscience.com/chemistry/plos-one-require-public-access-data.

First received 23 February 2015; accepted in revised form 23 October 2015. Available online 27 November 2015