

Publishing Data with CUAHSI Water Data Services

The CUAHSI Hydrologic Information System (HIS) is designed to improve access to the nation's water data. An important part of this information are time series of observations made at point locations, such as precipitation and streamflow gages, soil water and climate stations, groundwater wells and water quality sampling sites in surface and groundwater. These data can be stored in the CUAHSI Observations Data Model, communicated through the internet using the WaterML language, and cataloged in a national water metadatabase. Individual researchers and research organizations can use these facilities to publish their water data as a CUAHSI Water Data Service.

This document provides the interim data policies for data use and for data publication with CUAHSI Water Data Services. These policies are subject to review and to revision. Any changes to data use policies will be posted to the CUAHSI web site. Any changes to data publication policy will be distributed to all those who have published their data and subject to these policies. Comments on these policies are welcome and may be sent to exdir@cuahsi.org.

Guiding Principles

CUAHSI exists to serve the academic research community. Core corporate values are objectivity, transparency, and cooperation. CUAHSI WDS contribute towards achieving the goals of the hydrologic science community by making data more readily available and by enabling academic researchers to share data that they have collected. Furthermore, CUAHSI WDS strive to meet the requirements of NSF's data publication policy, so that data providers who choose to publish using WDS meet all NSF requirements.

Data users outside the academic research community will benefit from CUAHSI WDS and CUAHSI encourages others to use these services. These principles, however, will guide the allocation of limited resources of the CUAHSI WDS project.

Modes of Data Publication

There are three steps to the establishment of a CUAHSI Water Data Service:

- (1) Storing observations data in the CUAHSI Observations Data Model (ODM) ;
- (2) Providing access to these data through WaterOneFlow web services;
- (3) Indexing the resulting water data service at HIS Central in the national water metadata catalog.

For more information about these steps, see <http://his.cuahsi.org>

Any organization or individual is free to using CUAHSI Hydrologic Information System tools to accomplish the first two of these steps, and thus to establish their own water data services. This document

applies mainly to the third step in this process whereby the water data service becomes part of CUAHSI national inventory of water data services.

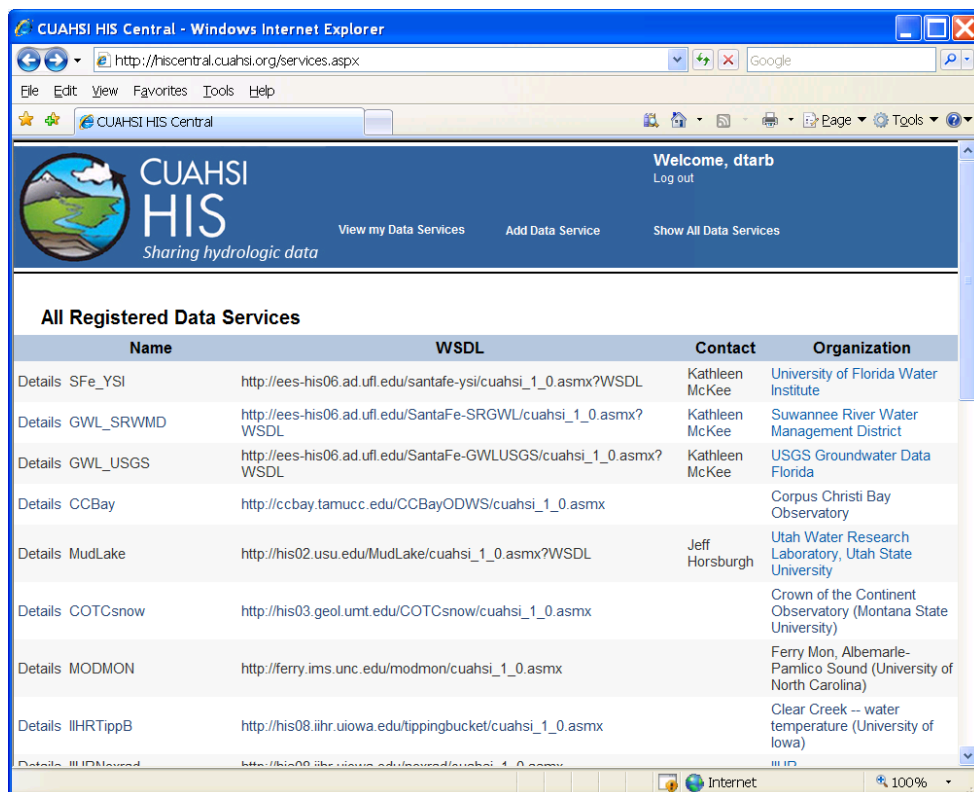
The database schema (Observations Data Model or ODM), Water Markup Language (WML), the web services developed for ODM, and all tools developed for the ODM (e.g., data loaders and editing tools) are freely distributed subject to the terms of the BSD License (<http://www.opensource.org/licenses/bsd-license.php>).

Data Service Registration

CUAHSI will maintain a website (<http://hiscentral.cuahsi.org/>) for the receipt of Web Service registration requests. This website will maintain a list of registered Web Service URLs and brief description of the content of the data service. The listing page and an example network page are shown below.

CUAHSI will maintain a system that populates these pages with user responses gathered during the registration process and by using web services to query the published data base. Those fields, such as number of variables, will be updated weekly by making calls to the registered WSDL address. If these calls fail to return a result, an e-mail will be sent to the registered data manager for the data service and a note will be placed on the network web page that the web service was down when last tested. Four successive failures will result in placement of the data service on inactive status. Data publishers may request de-registration from the service at any time.

CUAHSI will maintain this registration service so long as it has resources to do so. If, at the end of the current project, additional funds are not available for maintaining the page, the service will be suspended and all data publishers notified. The current CUAHSI HIS project is funded through December 31, 2011.



The screenshot shows the CUAHSI HIS Central website in a Windows Internet Explorer browser. The page has a blue header with the CUAHSI logo and the text "CUAHSI HIS Sharing hydrologic data". Below the header, there are links for "View my Data Services", "Add Data Service", and "Show All Data Services". The main content area is titled "All Registered Data Services" and contains a table with the following data:

Name	WSDL	Contact	Organization
Details SFe_YSI	http://ees-his06.ad.ufl.edu/santafe-ysi/cuahsi_1_0.asmx?WSDL	Kathleen McKee	University of Florida Water Institute
Details GWL_SRWMD	http://ees-his06.ad.ufl.edu/SantaFe-SRGWL/cuahsi_1_0.asmx?WSDL	Kathleen McKee	Suwannee River Water Management District
Details GWL_USGS	http://ees-his06.ad.ufl.edu/SantaFe-GWLUSGS/cuahsi_1_0.asmx?WSDL	Kathleen McKee	USGS Groundwater Data Florida
Details CCBay	http://ccbay.tamucc.edu/CCBayODWS/cuahsi_1_0.asmx		Corpus Christi Bay Observatory
Details MudLake	http://his02.usu.edu/MudLake/cuahsi_1_0.asmx?WSDL	Jeff Horsburgh	Utah Water Research Laboratory, Utah State University
Details COTCsnow	http://his03.geol.umt.edu/COTCsnow/cuahsi_1_0.asmx		Crown of the Continent Observatory (Montana State University)
Details MODMON	http://ferry.ims.unc.edu/modmon/cuahsi_1_0.asmx		Ferry Mon, Albemarle-Pamlico Sound (University of North Carolina)
Details IIHRTippB	http://his08.ihr.uiowa.edu/tippingbucket/cuahsi_1_0.asmx		Clear Creek -- water temperature (University of Iowa)
Details IIHRTippB	http://his08.ihr.uiowa.edu/tippingbucket/cuahsi_1_0.asmx		Clear Creek -- water temperature (University of Iowa)

Figure 1. Listing of WSDL's.

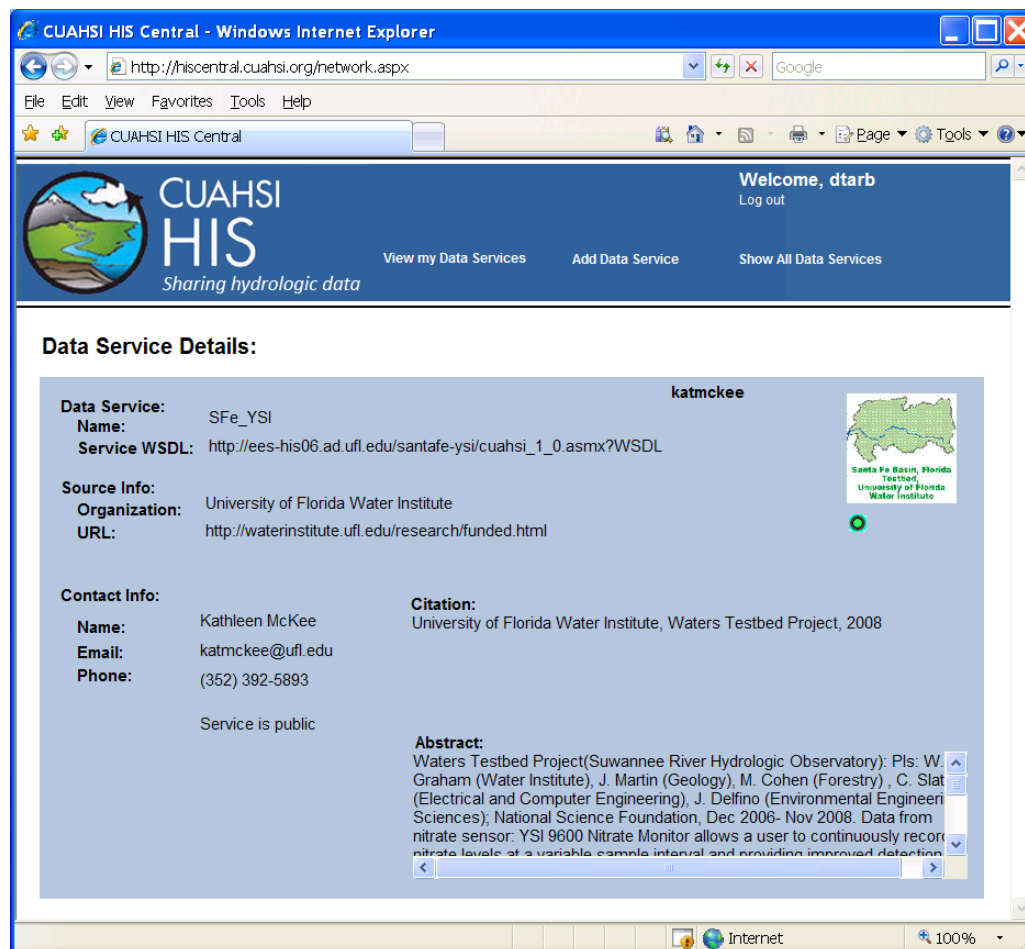


Figure 2. Description of data network.

User's Responsibilities

1. Acceptance of data publishing agreement (on-line, as part of registration process).
2. Maintenance web service on-line

CUAHSI's Responsibilities

1. Maintenance of registration service
2. Maintenance of web pages with weekly updates of catalog information obtained through web services
3. Notification of data manager of service outage

Data Indexing

Beyond simple registration of the WSDL address, CUAHSI also maintains a metadata catalog that enables discovery of data through the use of tools such as HydroSeek. In addition to the registration process, the data published must use HydroTagger (<http://his.cuahsi.org/hydrotagger.html>) to link

variables in the data base with concepts in the Water Resources ontology that CUAHSI has developed. Some variables will not be included in the concepts contained in the ontology. HydroTagger enables users to request a new concept be added. The ontology will be updated roughly every six months, so that there may be some variables that are not discoverable through the Data Indexing service. Note, however, that these variables can still be accessed using WDS.

To be eligible for data indexing services, various features of ODM must be respected. First, all required metadata fields must be entered. This is consistent with the policy that, while data quality standards cannot be stated a priori, documentation of data collection methods must be done to meaningfully communicate scientific data. Second, the controlled vocabulary contained within the ODM must be respected to maintain homogeneity across data services. If a site would like to include terms not contained in the CV, a request for additional terms, or edits to existing terms may be made at <http://his.cuahsi.org/mastercvreg.html>.

CUAHSI will update the metadata catalog weekly and maintain HydroSeek at production levels through the San Diego Supercomputing Center, at least for the duration of the current project. If the data service fails to respond to this weekly request, the data manager will be notified. If four successive weekly requests fail, the data service will be placed on an 'inactive' status. Sites may be re-activated by requesting re-activation at <http://hiscentral.cuahsi.org>.

CUAHSI will publish on at <http://hiscentral.cuahsi.org> quarterly reports of data download requests for each data publisher. Reporting web service hits and downloads is a part of our obligations to NSF. With each data download, the data user will be instructed how to cite the data, based upon information provided by the data user. Data publishers who use generic ODM web services to publish their data (as part of HIS Server or HIS Server Lite) can take advantage of the remote web service logging implemented in the services, and have the usage statistics summarized and published on the CUAHSI HIS site as described above (note that port 8090 shall remain open for this functionality to work). Alternately, CUAHSI is requesting that web service requests are logged and a summary is provided to CUAHSI quarterly. If a data publisher chooses not to implement a web service usage logging for his or her data, then CUAHSI could only track web service usage and data download requests through Hydroseek.

Because resources are required to maintain the metadata catalog, CUAHSI reserves the right to limit the number of data services it will index. During the current project period, resources will be allocated based upon the following priorities:

1. CUAHSI Members, Affiliates, and International Affiliates
2. Other US universities
3. Other US research institutes and non-profits
4. International universities
5. International non-profits
6. Others

Note that this prioritization applies to data providers who request data indexing services; resources for indexing Federal, State and Local data are decided separately based upon the value of the data to the academic research community.

User's Responsibilities

1. Acceptance of data publishing agreement (on-line, as part of registration process).
2. Maintenance web service on-line
3. Adherence to ODM requirements for metadata and controlled vocabulary
4. One-time linking of variables with concepts in CUAHSI Water Resources Ontology

CUAHSI's Responsibilities

1. Maintenance of registration service
2. Maintenance of web pages with weekly updates of catalog information obtained through web services
3. Notification of data manager of service outage
4. Maintenance of ontology and associated tools
5. Maintenance of Hydroseek discovery tool

Data Hosting

In those instances when a data publisher does not wish to maintain a server for publishing data, data sets may be loaded into an ODM, variable names tagged to concepts, and transmitted to SDSC for data publication and archiving. See <http://his.cuahsi.org> for details on the procedures for doing this. This service is the most resource-intensive of all the WDS offered. CUAHSI and SDSC reserve the right to limit the number of data sets that will be hosted. The same prioritization applies as with Data Indexing. This hosting service will be maintained at least through the end of the current project (Dec. 31, 2011). Plans for an on-going data center to provide this service are under development.

Users may request that their data sets be deleted from the hosting service at any time by a request filed at <http://hiscentral.cuahsi.org>.

User's Responsibilities

1. Acceptance of data publishing agreement (on-line, as part of registration process).
2. Maintenance web service on-line
3. Adherence to ODM requirements for metadata and controlled vocabulary
4. One-time linking of variables with concepts in CUAHSI Water Resources Ontology.
5. Verification that data were transmitted faithfully to SDSC

CUAHSI's Responsibilities

1. Maintenance of registration service
2. Maintenance of web pages with weekly updates of catalog information obtained through web services
3. Notification of data manager of service outage
4. Maintenance of ontology and associated tools
5. Maintenance of Hydroseek discovery tool
6. Maintenance of data set on production server

CUAHSI Water Data Services Data Use Agreement

This document outlines the provisions of a non-exclusive license for use of data shared through the CUAHSI Water Data Services. Consistent with the objectives of CUAHSI, the goal of the CUAHSI WDS is to make data originally acquired with investigator-specific interest available to the community for further study. All data are freely available to the public. All Users are encouraged to consider sharing resultant data through CUAHSI WDS to extend and strengthen this collaborative environment.

By receipt and use of this data, you agree to the following provisions for yourself *and any collaborators with whom you share these data*:

1. Data Guarantee:

I will employ these data at my own risk, as the quality of these data cannot be guaranteed.

2. Publication / Acknowledgement of Data Use:

Acknowledgement of the data provider(s) is expected as standard practice in scientific publication or presentation of findings based upon these data. Whenever practical, the individual data providers should be acknowledged; when impractical to do so because of the number of sources of data, the use of CUAHSI WDS data should be acknowledged.

I agree to provide to CUAHSI a bibliographic citation of the final published presentation or article for inclusion in the CUAHSI literature archive. I will send this information to **exdir@cuahsi.org**.

3. Redistribution of Data:

Redistribution of original data is permitted so long as the data are redistributed under the same terms and conditions as described in this Data Use Agreement.

Data derived from original data may be distributed under terms and conditions established by the creators of such derived data. Users must comply with the terms and conditions of use set by the creators of these derived data.

CUAHSI Water Data Services Data Publisher's Agreement

To support a collaborative research environment, CUAHSI WDS publishes data generated by the hydrologic sciences research community and makes these data freely available. CUAHSI Data publication services also support fulfillment of the data sharing policy of the National Science Foundation's Earth Sciences Division (http://www.nsf.gov/geo/ear/EAR_data_policy_204.pdf). Your participation will encourage scientific inquiry, enable new research exploration, and facilitate education by providing the scientific community with relevant, easily accessible data. Shared data are available to the public according to the standard CUAHSI WDS Data Use Agreement.

By your acceptance of these terms, you state that you are the owner of the data or have the right to publish the data. You further agree to the following provisions for yourself *and any collaborators with whom you coordinate the submission of these data*:

1. As the Contributor, I am solely responsible for the integrity of these data put forward for submission.
2. I will submit a data publication request to the CUAHSI Web Service registry, giving details pertaining to the dataset being contributed.
3. I will maintain a current contact information, including name, organization, phone number, e-mail address, and acceptance of a standard CUAHSI WDS Data Use Agreement at the CUAHSI Web Service registry (<http://hiscentral.cuahsi.org>).
4. I will provide a high-level description of the study through which the data were originally acquired.
5. I will provide a minimum set of associated metadata to sufficiently describe submitted data, to ensure that shared data are meaningful and useful to the scientific community.
6. I agree to work with the CUAHSI Hydrologic Information Systems project staff to ensure that these data are provided in readable formats.
7. I agree to make these data publicly available in accordance with an agreed upon timeline with no restrictions for use other than those stated in the standard CUAHSI WDS Data Use Agreement.
8. I understand that upon registration at the CUAHSI Web Service registry, CUAHSI will retrieve and index metadata describing my dataset (information about sites, variables, periods of record) to support rapid data discovery via CUAHSI Hydrologic Metadata Catalog.
9. CUAHSI is interested in reporting web service usage statistics and data downloads to NSF. For this purpose it maintains a remote web service logging infrastructure which ensures that usage statistics is submitted to a central logging system. As a Data Contributor, I can choose whether to use the logging system and take advantage of uniform quarterly reporting of web service hits and data downloads, or maintain my own usage statistics and provide quarterly updates to CUAHSI. The components of the logging infrastructure are part of the Generic ODM web services, and require port 8090 to be open.