



CUAHSI
UNIVERSITIES ALLIED FOR WATER RESEARCH

A USER'S GUIDE TO THE CLOUD HYDROSERVER

**Cloud Publication
with the
CUAHSI Water Data Center**

Version 1.2

**Prepared by
Liza Brazil, Community Support Specialist
Jon Pollak, Program Manager**

July 2016

CONTENTS

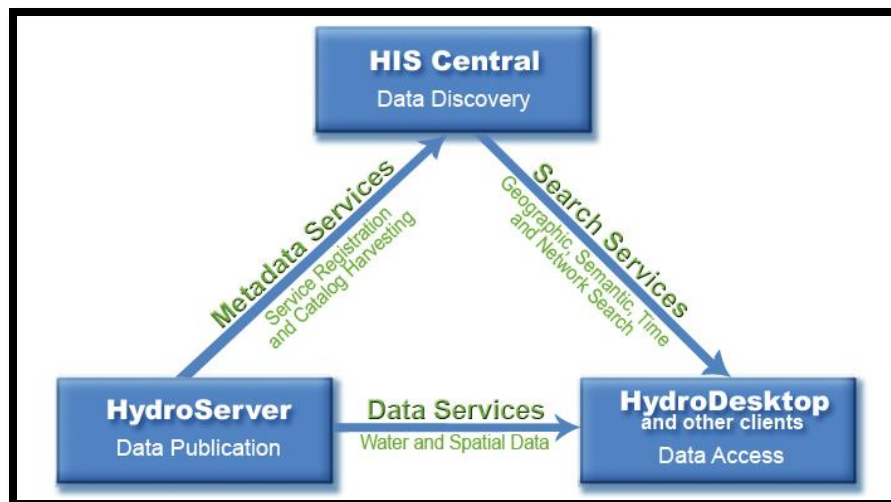
Disclaimer.....	3
Introduction and Purpose of This Document.....	4
Workflow for Publishing Data in CUAHSI HIS With Cloud Services	4
User request.....	5
Approval and Notification by CUAHSI WDC Staff	5
Populating The Database	5
Formatting Data	5
Uploading Data	6
Viewing and Editing Data in the Database.....	9
Registering a Data Source with HIS Central	10
Authenticating on HIS Central.....	11
Adding a Data Service and Metadata.....	11
The HIS Central Metadata Harvester	14
Appendix A: The Observations Data Model (ODM)	15
Appendix B: Required and Optional Templates.....	16
Appendix C: Controlled Vocabularies in the ODM.....	17

DISCLAIMER

The CUAHSI Water Data Center (WDC) is a facility funded by the National Science Foundation to enable access to water data holdings, create and maintain data discovery tools, and provide archival and data publishing capabilities. The WDC is funded by National Science Foundation grant number 1248152. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

INTRODUCTION AND PURPOSE OF THIS DOCUMENT

The CUAHSI Hydrologic Information System has been developed in order to enable the sharing of water data between researchers and groups located anywhere in the world. It is composed of three parts. The first component is a data server (HydroServer), which holds and provides data upon request. The second is the metadata catalog (HIS Central), which facilitates search using metadata. The last component is the client used to search and retrieve data (e.g. HydroClient).



Above: The three components and linkages of the CUAHSI Hydrologic Information System.

Web services, or computer programs that request and transmit data, facilitate the interaction between these three components.

The focus of this document is on the data storage component of the CUAHSI HIS, HydroServer, which has been deployed in the Microsoft Azure Cloud. The content of this document describes how to request using these cloud services from CUAHSI, how to populate a database in the cloud, and how to register as a data source on the CUAHSI Water Data Center's central metadata catalog, HIS Central.

The interface for CUAHSI's cloud-based HydroServer is located on the domain <http://hydroportal.cuahsi.org>.

WORKFLOW FOR PUBLISHING DATA IN CUAHSI HIS WITH CLOUD SERVICES

There are four general steps for publishing data in the cloud with the CUAHSI WDC:

1. Requesting a Database
2. Review by CUAHSI WDC Staff
3. Population of the Database

4. Registering the Database as a Data Source in HIS Central

USER REQUEST

Access to the CUAHSI WDC's hosting services is limited to approved data publishers. Requests for databases can be sent to help@cuahsi.org and must include the following information:

- Name of person requesting the database
- Affiliation of person requesting the database
- Contact email
- Google account that will be used for authentication in the WDC system¹
- Proposed name of network of sites (A unique code associated with the data source)

APPROVAL AND NOTIFICATION BY CUAHSI WDC STAFF

Once a request has been approved, CUAHSI WDC staff will take the following actions:

1. Deploy an instance of the ODM Database
2. Deploy an instance of WaterOneFlow web services
3. Configure the ODM database and WaterOneFlow web service so that they are linked to the uploading application and authenticated through the user's Google account
4. Notify the user when the above items have been successfully provisioned and provide user with URL for data service

POPULATING THE DATABASE


Populating the database occurs in two steps:

1. Formatting data to comply with the Observations Data Model (ODM) using CUAHSI provided templates
2. Uploading templates that contain metadata and data

FORMATTING DATA

The ODM Uploader for the cloud contains templates in the form of .csv files under the Templates tab of the website, which are seen in the screenshot below.

¹ This can either be a .Gmail domain or another domain that is using Google's authentication protocol (such as an institution that uses Google Apps as its email provider).



HydroServer

CUAHSI Open Source Hydrologic Data Tools

[Home](#)
[Contact](#)
[Manage](#)
[Templates](#)
[Controlled Vocabularies](#)
[Import CSV Data ▾](#)
[View Data ▾](#)
[Admin ▾](#)
Hello lizabrazil505@gmail.com!

Templates

Please click to download the templates in CSV Format

- [Sites](#)
- [Variables](#)
- [OffsetTypes](#)
- [Sources](#)
- [Methods](#)
- [LabMethods](#)
- [Samples](#)
- [Qualifiers](#)
- [QualityControlLevels](#)
- [DataValues](#)
- [GroupDescriptions](#)
- [Groups](#)
- [DerivedFrom](#)
- [Categories](#)

Please click to download User Guide in PDF Format

- [UserGuide](#)

Please click to download User Data Upload Tutorial in PDF Format

- [Data Upload Tutorial](#)

Please click to download ODM Guide in Excel (xlsx) Format

- [ODMGuide](#)

Please click to download ODM Guide in Open Office Format

- [ODMGuide](#)

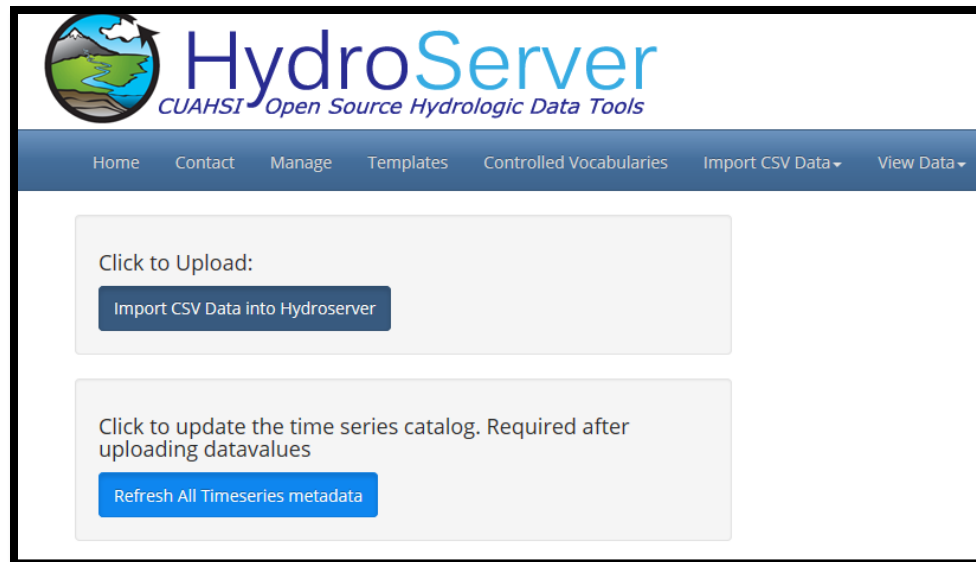
[Report a bug here!](#)

© 2016 - © CUAHSI Version 1.1.5

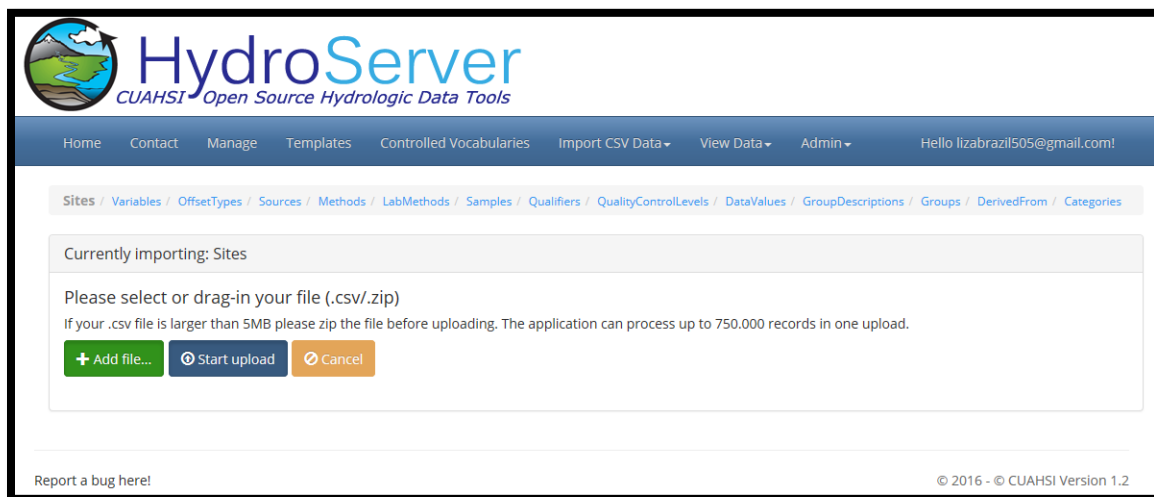
In the left column, there are templates that represent all of the tables that comprise the ODM. Not all of these tables are mandatory, however. For a detailed look at what tables and fields are required, there is an ODM Guide in .xlsx format also located on the Templates page on the Cloud HydroServer website, which is located on the right side of the page. This file contains a color-coded explanation of mandatory and optional tables, mandatory and optional fields, as well as controlled vocabulary fields. This document was derived from the technical specifications for the ODM.

UPLOADING DATA

To begin uploading templates either click on Import CSV Data into HydroServer on the landing page or click the dropdown for Import CSV Data.



The workflow for uploading templates is an ordered process. Although not all of the templates are required, the upload page for each is listed. From left to right: **Sites**, **Variables**, **OffsetTypes**, **Sources**, **Methods**, **LabMethods**, **Samples**, **Qualifiers**, **QualityControlLevels**, **DataValues**, **GroupDescriptions**, **Groups**, **DerivedFrom**, and **Categories**.



Templates that have been populated with data can be uploaded by going to the corresponding upload page for the template. The HydroServer can only handle .csv files that contain less than 5MB, for anything larger you must compress the .csv into a .zip file. Begin an upload by either dragging a .csv or compressed .zip file on to the *Add File...* button or click *Add File...* and navigate to the location of the compressed .zip file. Clicking *Start Upload* will upload the data to CUAHSI's server for validation.

The application will return a webpage that appears similar to the screenshot below where the rows from your .csv file have been parsed into four categories:

- **New:** Valid records ready to be committed to the database.
- **Rejected:** Records that could not be validated by the application.

- **Updated:** Records that have been updated because the uploaded record had a previously used key with other, changed metadata that are ready to be committed to the database.
- **Duplicate:** Records that are a complete match of records that have been previously committed to the database.

The screenshot shows the HydroServer web application interface. The top navigation bar includes links for Home, Contact, Manage, Templates, Controlled Vocabularies, Import CSV Data, View Data, Admin, and a user greeting. Below the navigation bar is a breadcrumb trail. The main content area features a tabbed interface with four tabs: New (0), Rejected (4), Updated (0), and Duplicate (0). The 'Rejected' tab is active, displaying a table of rejected records. The table has columns: DataValue, ValueAccuracy, LocalDateTime, UTCOffset, DateTimeUTC, SiteCode, and Vari. The first row shows a rejected record with a red minus icon in the far left column. Below the table, there is a message: 'The value 1127000 is not in SiteCode Table. Please validate your input.' Below this message is a 'Download as CSV' button. At the bottom of the interface, there is a 'Commit Changes' button and an 'Update Timeseries metadata' button.

	DataValue	ValueAccuracy	LocalDateTime	UTCOffset	DateTimeUTC	SiteCode	Vari
✖	0		1/1/1948	0	1/1/1948	1127000	
The value 1127000 is not in SiteCode Table. Please validate your input.							
✔	-3.6722		1/3/1948	0	1/3/1948	1127000	
✔	1		1/4/1948	0	1/4/1948	1127000	
✔	1		1/2/1948	0	1/2/1948	11270005	

Showing 1 to 4 of 4 entries

First Previous 1 Next Last

✖ Cancel Download as CSV

Click "Update Timeseries metadata" after data is saved to the database. This is required to register the new data values with the existing timeseries or create new timeseries. Failing to do this will result in the new data not being discoverable. Since this can take a long period of time depending on the amount of data in the database it is recommended to be the last step of uploading multiple sheets of data.

Update Timeseries metadata

Navigate between the four categories by clicking the labeled tabs. In the *Rejected* tab, view error messages by clicking the green plus symbol in the far left column. Download the rejected records as a .csv file by clicking the *Download as CSV* button on the right as seen in the image above.

To commit the new and updated records to the database click the *Commit Changes* button, which is located in the lower left of the *New* tab. When uploading new Data Values, you'll need to click *Update Timeseries metadata* button. If you are uploading a batch of new data values, you may wait to click *Update Timeseries metadata* once all data values have been committed, as this process can take several minutes to complete.

New **11**
Rejected **2**
Updated **0**
Duplicate **0**

Show **10** entries
Search:

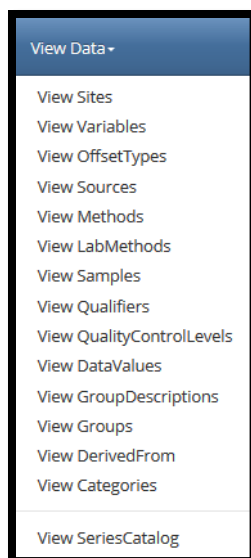
	VariableCode	VariableName	Speciation	VariableUnits Name	SampleMedium	ValueType	IsRegular	TimeSupport
	Ammonium	Nitrogen, NH4	Unknown	micron	Surface water	Sample	TRUE	0
	chloro_a	Chlorophyll a	Unknown	micrograms per liter	Surface water	Sample	TRUE	0
	particulate_carbon	Carbon, particulate organic	Unknown	micron	Surface water	Sample	TRUE	0
	Particulate_nitrogen	Nitrogen, particulate organic	Unknown	micron	Surface water	Sample	TRUE	0
	particulate_phosphorus	Phosphorus, particulate	Unknown	micron	Surface water	Sample	TRUE	0
	Pheophytin	Pheophytin	Unknown	micrograms per liter	Surface water	Sample	TRUE	0
	phosphate	Phosphorus, phosphate (PO4)	Unknown	micron	Surface water	Sample	TRUE	0
	Total_dissolved_nitrogen	Nitrogen, total dissolved	Unknown	micron	Surface water	Sample	TRUE	0
	Total_dissolved_phosphorus	Phosphorus, total dissolved	Unknown	micron	Surface water	Sample	TRUE	0
	total_kjeldahl_nitrogen	Nitrogen, dissolved Kjeldahl	Unknown	micron	Surface water	Sample	TRUE	0

Showing 1 to 10 of 11 entries
First Previous **1** 2 Next Last

Commit Changes
Cancel
Refresh All Timeseries metadata
Download as CSV

VIEWING AND EDITING DATA IN THE DATABASE

To view data that have already been loaded into the database, click the View Data and select the table that you wish to view.



The resulting webpage will allow you to view, search, and sort through data that have been committed to the database.

HydroServer
CUAHSI Open Source Hydrologic Data Tools

Home Contact Manage Templates Controlled Vocabularies Import CSV Data View Data Admin Hello lizabrazil505@gmail.com

Sites

Show 10 entries Search:

Site Code	Site Name	Latitude	Longitude	LatLongDatum	RSName	Elevation_m	VerticalDatum	LocalX
105	Outer Harbor 105	42.326299	-70.948951	Unknown				
106	Outer Harbor 106	42.330614	-70.966117	Unknown				
117	Hingham Bay 117	42.291261	-70.916335	Unknown				
118	Outer Harbor 118	42.331629	-70.936248	Unknown				
124	Hingham Bay 124	42.275767	-70.899513	Unknown				
129	Winthrop Bay 129	42.356751	-70.98088	Unknown				
130	Winthrop Bay 130	42.365376	-70.991008	Unknown				
137	Mystic mouth 137	42.386934	-71.065337	Unknown				
138	Mid-inner harbor 138	42.361063	-71.040962	Unknown				
139	Quincy Bay 139	42.280086	-70.970237	Unknown				

Showing 1 to 10 of 46 entries

First Previous 1 2 3 4 5 Next Last

Report a bug here! © 2016 - © CUAHSI Version 1.1.5

Editing the contents of a database requires a new upload. For Sites and Variables, uploading a record with a previously used Code (SiteCode or VariableCode) will update the corresponding record in the database with the information from the new upload. Such records will be presented in the validation step during upload under the *Updated* tab. Other edits must be performed by first deleting an entire table, in the Manage page, then uploading new records.

REGISTERING A DATA SOURCE WITH HIS CENTRAL

Registering your data service as a data source in the WDC's HIS Central Catalog requires just two items:

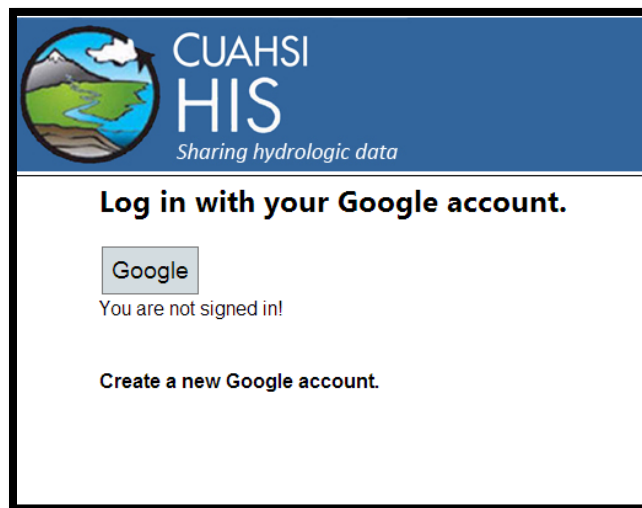
1. A Google account for authentication
2. The URL for the data service (provided by CUAHSI staff)

The following tasks must be completed in order for data to be discoverable through data access software like HydroClient:

1. Authenticate on the HIS Central website (hiscentral.cuahsi.org)
2. Add URL for the data service and applicable metadata
3. Await or request a harvest of the metadata

AUTHENTICATING ON HIS CENTRAL


The easiest way to sign on to HIS Central, like the Cloud HydroServer website, is by using a Google account.



To sign on to HIS Central, visit the website (<http://hiscentral.cuahsi.org>), click Login, and follow the prompts to allow HIS Central to access your Google account.

ADDING A DATA SERVICE AND METADATA

Once signed on to the HIS Central website, it is then possible to add a data service to the catalog. To do this, click Add Data Service, give the data source a Service Title, fill in the Network Name that was used in the database request to CUAHSI, and finally fill in the WSDL URL, which is provided by CUAHSI staff.



**CUAHSI
HIS**
 Sharing hydrologic data

Welcome, jtpollak@gmail.com
 Log out
 Administration

[Home](#)
[My Data Services](#)
[Add Data Service](#)
[All Data Services](#)
[Forum](#)

Register Data Service


Service Title:
This will appear atop the page associated with your service.

Network Name:
This is a unique code associated with your service. The network name used when configuring your webservice is appropriate. This value is unique across this system.

Service WSDL:
Enter the full web URL to your webservice WSDL. This value is unique across this system.

☐ [I have read and agree to the Data Service Agreement](#)

After a new data service is added to the HIS Central catalog, the next step is to add additional metadata about the publishing entity. This can be done by clicking My Data Services tab, then clicking the Details link for a specific service, which can be seen on the left side of the screenshot below.



**CUAHSI
HIS**
 Sharing hydrologic data


Welcome, jtpollak@gmail.com
 Log out
 Administration

[Home](#)
[My Data Services](#)
[Add Data Service](#)
[All Data Services](#)
[Forum](#)

My Services:

		NetworkName	ServiceWSDL
Details	Test Usage Report	MuddyRiver	http://hydroportal.cuahsi.org/muddyriver/cuahsi_1_0.asmx?WSDL
Details	Test Usage Report	GlacialRidge	http://hydroportal.cuahsi.org/glacialridge/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	LaSelvastreamdischarge	http://hydroportal.cuahsi.org/LaSelvastreamdischarge/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	TuolumneMdw	http://hydroportal.cuahsi.org/TuolumneMdw/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	Hassberge	http://hydroportal.cuahsi.org/hassberge/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	Weiherbach	http://hydroportal.cuahsi.org/weiherbach/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	IRWA	http://hydroportal.cuahsi.org/ipswich/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	IEER at Wilkes University	http://hydroportal.cuahsi.org/Wilkes/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	Panola ODM	http://hydroportal.cuahsi.org/panolaemma/webapp/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	Swedish Monitoring Data	http://hydroportal.cuahsi.org/SwedishMonitoringData/webapp/cuahsi_1_1.asmx?WSDL
Details	Test Usage Report	UBWPAD	http://hydroportal.cuahsi.org/umassblackstone/webapp/cuahsi_1_1.asmx?WSDL

The following page will show a preview of the metadata for the data publisher. To edit this information, click Edit Details, which is seen in the lower left of the screenshot below.



**CUAHSI
HIS**
Sharing hydrologic data

Welcome, jtpollak@gmail.com
Log out
Administration

HomeMy Data ServicesAdd Data ServiceAll Data ServicesForum

Data Service Details: [View Public Details Page](#)

Data Service Title: Panola Chemistry Data for EMMA
Network Vocab: Panola ODM
Service WSDL: http://hydroportal.cuahsi.org/panolaemma/webapp/cuahsi_1_1.asmx?WSDL

Source Info:
Organization: United States Geological Survey
URL: <http://ga.water.usgs.gov/projects/panola/>

Contact Info:
Name: Rick Hooper
Email: RHooper@cuahsi.org
Phone: 339-221-5400

Service is public

LastHarvested: 6/7/2014 7:07:05 PM
Not Approved


Citation:
Christophersen, N. and Hooper, R.P., 1992. Multivariate analysis of stream water quality. The use of principal components for the end-member mixing problem. *Water Resour. Res.*, 28, 99-107.

Abstract:
These data were used by Hooper et al (1990) and Christophersen and Hooper (1992) in the initial development of End Member Mixing Analysis.


Edit Details

Change Images

Upload a custom logo and map icon to be used for your data service



The subsequent page will allow you to edit information such as contact information for the data publisher:



**CUAHSI
HIS**
Sharing hydrologic data

Welcome, jtpollak@gmail.com
Log out
Administration

HomeMy Data ServicesAdd Data ServiceAll Data ServicesForum

Data Service Details: [View Public Details Page](#)

Service Title: Panola Chemistry Data for EMMA
Network Vocab: Panola ODM
Service WSDL: http://hydroportal.cuahsi.org/panolaemma/webapp/cuahsi_1_1.asmx

Source Info:
Organization: United States Geological Survey
URL: <http://ga.water.usgs.gov/projects/panola/>
What organization is publishing this data?

Contact Info:
Name: Rick Hooper
Email: RHooper@cuahsi.org
Phone: 339-221-5400
Who is the primary contact?

☒ Is service public?
Service must be public to be accessible through this portal.


Citation:
Christophersen, N. and Hooper, R.P., 1992. Multivariate analysis of stream water quality: The use of principal components for the end-member mixing problem. *Water Resour. Res.*, 28, 99-107.
How do you want your data to be cited when downloaded?

Abstract:
These data were used by Hooper et al (1990) and Christophersen and Hooper (1992) in the initial development of End Member Mixing Analysis.
Provide a brief description about your project and how you collected your data.

UpdateCancel

Change Images

Upload a custom logo and map icon to be used for your data service

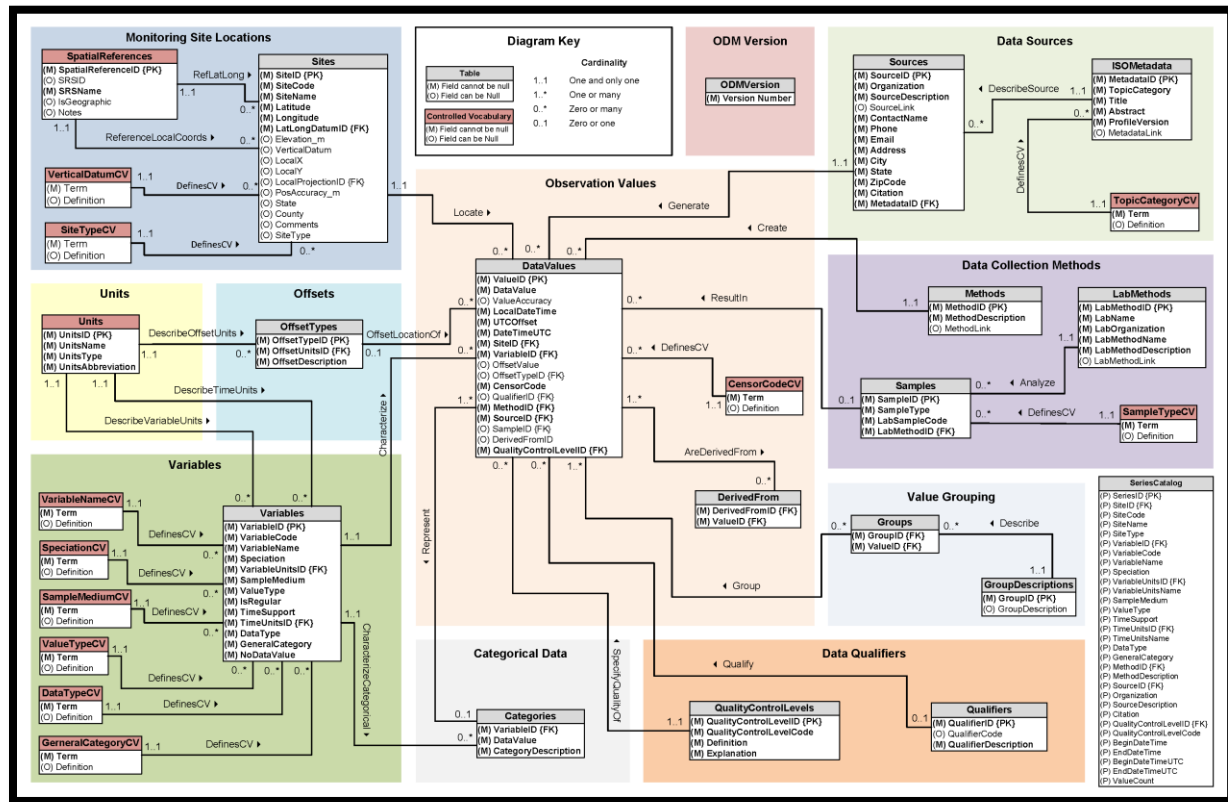


THE HIS CENTRAL METADATA HARVESTER

The HIS Central Catalog utilizes a harvester to gather metadata from each data source registered in the catalog. Much like a web crawler that visits and gathers data from websites, this harvester is a computer program that visits each registered data source and gathers metadata. The metadata collected and stored by the HIS Central enables the search and discovery of data. Data services that are registered are harvested once a week. If a data publisher needs their data source to be harvested prior to the next scheduled harvest they can request it from CUAHSI staff.

Once a data source has been harvested by HIS Central, it will be discoverable by clients that search the catalog like HydroDesktop.

APPENDIX A: THE OBSERVATIONS DATA MODEL (ODM)



The Observations Data Model (ODM) is the information model employed by the CUAHSI HIS. The most common implementation of it is as relational database in Microsoft SQL. The core of the model comprises of a center table that contains the value of observations as well as foreign keys to ancillary tables. These tables provide metadata with the goal of unambiguous interpretation of the data values and include tables with information related to the:

- Location of the observation (*Sites*)
- Phenomenon being observed (*Variables*)
- Methods being employed (*Methods*)
- Sources of the data (*Sources*)
- Quality control techniques employed (*Data Qualifiers*)

The ODM is described in a peer-reviewed article in [Water Resources Research](#) and can be accessed [here](#).

For additional information, [visit the WDC website](#).

APPENDIX B: REQUIRED AND OPTIONAL TEMPLATES

Template Name	Template Description
<i>Variables</i>	The Variables template lists the full descriptive information about what variables have been measured. This template is mandatory.
<i>Methods</i>	The Methods template lists the methods used to collect the data and any additional information about the method. This template is mandatory.
<i>Sites</i>	The Sites template provides information giving the spatial location at which data values have been collected. This template is mandatory.
<i>Sources</i>	The Sources template lists the original sources of the data, providing information sufficient to retrieve and reconstruct the data value from the original data files if necessary. This template is mandatory.
<i>Samples</i>	The Samples template gives information about physical samples analyzed in a laboratory. This template is optional.
<i>LabMethods</i>	The LabMethods template contains descriptions of the laboratory methods used to analyze physical samples for specific constituents. This template is optional.
<i>Quality Control</i>	The QualityControlLevels template contains the quality control levels that are used for versioning data within the database. This template is mandatory.
<i>DataValues</i>	The DataValues template contains the actual data values and keys to metadata templates. This template is mandatory.
<i>Categories</i>	The Categories template defines the categories for categorical variables. This template is mandatory when variables exist that have DataType specified as "Categorical." Multiple entries for each VariableCode, with different DataValues provide the mapping from DataValue to category description.
<i>DerivedFrom</i>	The DerivedFrom template contains the linkage between derived data values and the data values that they were derived from. This template is optional.
<i>GroupDescriptions</i>	The GroupDescriptions template lists the descriptions for each of the groups of data values that have been formed. This template is optional and only required if the Groups template is used.
<i>Groups</i>	The Groups template lists the groups of data values that have been created and the data values that are within each group. This template is optional.
<i>Qualifiers</i>	The Qualifiers template contains data qualifying comments that accompany the data. This template is optional.

APPENDIX C: CONTROLLED VOCABULARIES IN THE ODM

Within the ODM, certain fields have been designed as controlled vocabularies, which are community driven and moderated by CUAHSI staff. Click here to visit the landing page for the [Master Controlled Vocabulary Registry for ODM 1.1](#).

The list of fields that are controlled vocabularies can be seen below listed by template and hyperlinked to the appropriate web page to view, submit edits, or submit new terms:

Variables

- [VariableName](#)
- [Speciation](#)
- [VariableUnitsName](#)
- [SampleMedium](#)
- [ValueType](#)
- [TimeUnitsName](#)
- [DataType](#)
- [GeneralCategory](#)

Sites

- [LatLongSRSName](#)
- [LocalProjectionSRSName](#)

Sources

- [TopicCategory](#)

Samples

- [SampleType](#)

DataValues

- [CensorCode](#)