ADVANCED UPLOADING GUIDE

Time Series and Sensor Data Publication on data.cuahsi.org

Version 2.0

Prepared by
Liza Brazil, Product Manager
Jon Pollak, Program Manager

March 2018

CONTENTS

DISCLAIMER

INTRODUCTION AND PURPOSE OF THIS DOCUMENT

The CUAHSI Hydrologic Information System (HIS) enables data uploading (HydroServer) and data discovery (HydroClient). This document describes how to create a publishing account, upload data, and request publication to make your data accessible to the public on HydroClient. Before proceeding with uploading your data, you must format it to fit with one of the two CUAHSI formatting options: The Standard Format Option or The Advanced Format Option. The Standard Format Option is a subset of the Advanced Format Option. Instructions on how to format your data and which format option to select are provided in the CUAHSI Formatting Guide.

The HydroServer provides two uploading options that correspond to the two options for formatting data. The Standard Upload option involves uploading only the required tables: Variables, Sites, Sources, Methods, Quality Control Levels, and Data Values. Refer to the Advanced Uploading Guide for instructions on uploading the full metadata profile, including all required tables and any optional metadata tables, e.g. Samples, LabMethods, Categories, Qualifiers.

WORKFLOW FOR PUBLISHING DATA

How to publish data in CUAHSI HIS:
1. Create a publishing account
2. Verify your account using the activation email
3. Format your data: Follow the Guide and fill in the Excel Templates (Standard or Advanced)
4. Upload your data: Follow the Guide (Standard or Advanced) to complete the process
5. Request Publication: Click Request Publication on the HydroServer Homepage to make your data service accessible to the public on http://data.cuahsi.org
6. For all subsequent uploads, data is published every Saturday automatically unless requested otherwise

Questions? Contact CUAHSI support at help@cuahsi.org

CREATE PUBLISHING ACCOUNT

You will need a Google Account to create a publishing account. The Google Account you use when creating a publishing account will be associated with the HydroServer. You may decide to create a shared Google Account for your organization or you may have several individual Google Accounts linked to one database.
To create a publishing account, go to http://hydroserver.cuahsi.org/ and click Create New Publishing Account. Fill out the required account information as described below.
- Service Name: An abbreviated name for your data service. Example: NWISDV
- Service Title: The full name of your data service. This will be displayed to the public. Example: NWIS Daily Values
- Service Abstract: A detailed description of your data service and the data.
- Contact Name: Name of the person that the public should contact with questions about the data.

- Contact Email: Email of the person that the public should contact with questions about the data.
- Contact Phone: Phone number of the person that the public should contact with questions about the data.
- Organization: Name of the organization responsible for publishing the data.
- Organization Website: Website address of the organization responsible for publishing the data.

Note: CUAHSI Staff will use the information from your publishing account to create an HIS Central Data Service page. This will list you among all data service providers in the CUAHSI Catalog. CUAHSI Staff will provide you with credentials to edit this information at any time.

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.
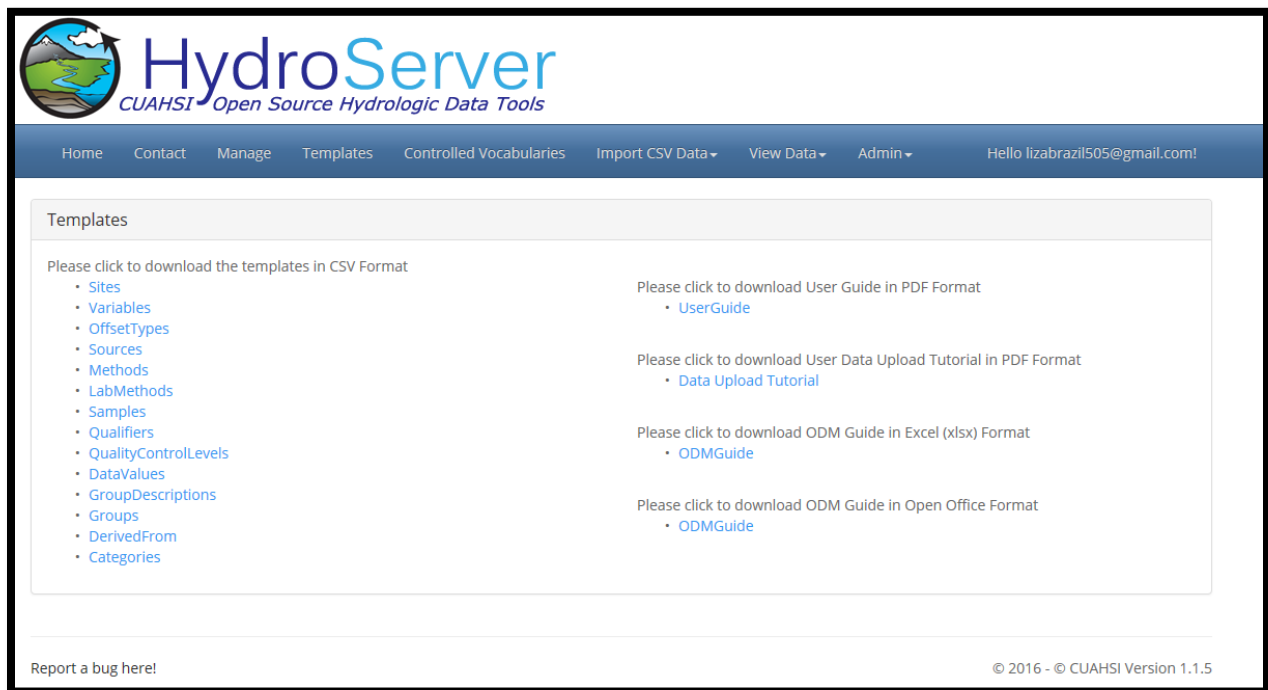

In the left column, there are templates that represent all of the tables that comprise the ODM. Not all of these tables are mandatory, however. For a detailed look at what tables and fields are required, there is an ODM Guide in .xslx format also located on the Templates page on the Cloud HydroServer 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 click on Advanced Upload on the HydroServer homepage. 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**.
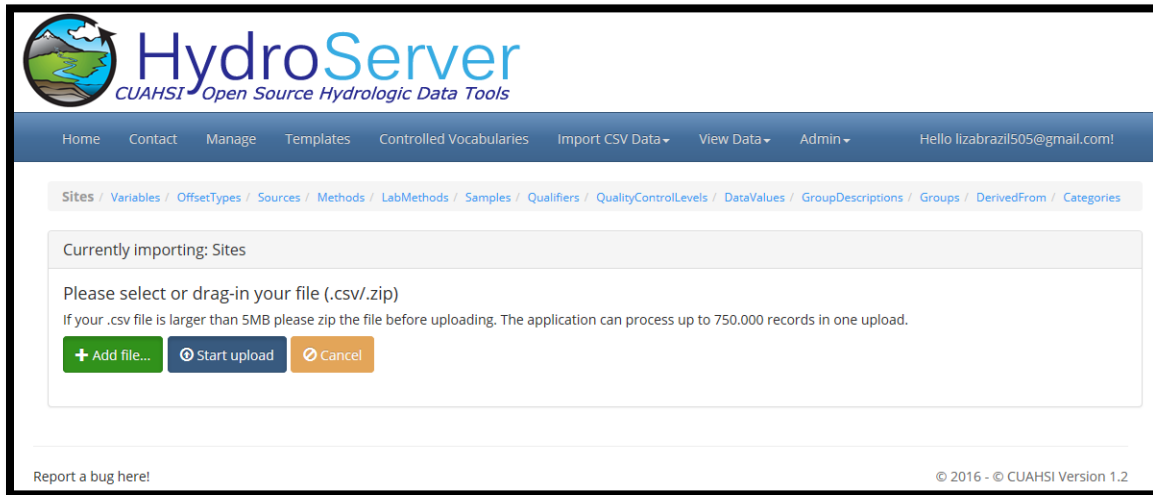
Templates that have been populated with data can be uploaded by going to the corresponding upload page for the template. The HydroServer can only handle .csv files that contain less than 5MB, for anything larger you must compress the .csv into a .zip file. Begin an upload by either dragging a .csv or 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.

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.

	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_car bon	Carbon, particulate organic	Unknown	micron	Surface water	Sample	TRUE	0
	Particulate_nit rogen	Nitrogen, particulate organic	Unknown	micron	Surface water	Sample	TRUE	0
	particulate_ph osphorus	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_dissolve d_nitrogen	Nitrogen, total dissolved	Unknown	micron	Surface water	Sample	TRUE	0
	Total_dissolve d_phosphorus	Phosphorus, total dissolved	Unknown	micron	Surface water	Sample	TRUE	0
	total_kjeldahl_ nitrogen	Nitrogen, dissolved Kjeldahl	Unknown	micron	Surface water	Sample	TRUE	0

New 11 Rejected 2 Updated 0 Duplicate 0

Show 10 entries Search: [   ]

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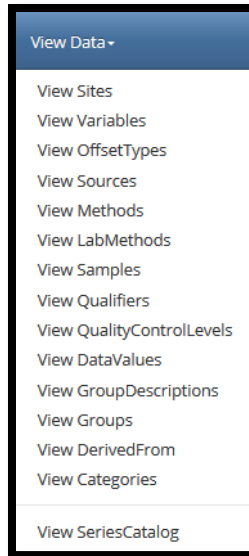
The resulting webpage will allow you to view, search, and sort through data that have been committed to the database.

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.

Once you are ready to publish your data to be discovered on http://data.cuahsi.org/, click **Request Publication** on the HydroServer homepage. A request will be sent to help@cuahsi.org to publish your data. Expect a response from help@cuahsi.org with a link to review your data on a private version of http://data.cuahsi.org/. Once you have reviewed and confirmed to help@cuahsi.org , CUAHSI staff will publish your data and create an HIS Central Data Service page that contains the information you provided when creating a publishing account. CUAHSI Staff will provide you with login information to access your HIS Central Data Service page and make any edits to the information.

Note: Data services are harvested once a week. If a data publisher needs their data to be harvested prior to the next scheduled harvest they can send a request to help@cuahsi.org.

APPENDIX A: THE OBSERVATIONS DATA MODEL (ODM)
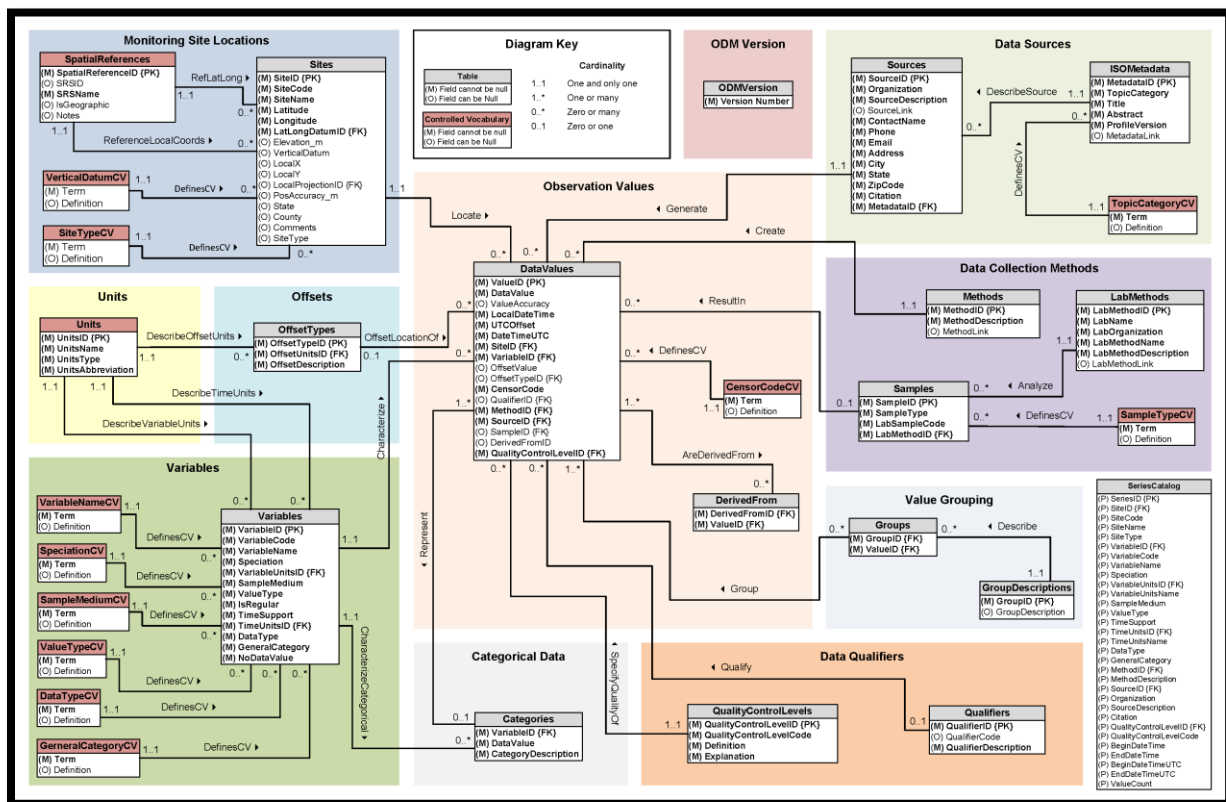
The Observations Data Model (ODM) is the information model employed by the CUAHSI HIS. The most common implementation of it is as relational database in Microsoft SQL. The core of the model comprises of a center table that contains the value of observations as well as foreign keys to ancillary tables. 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
Variables	The Variables template lists the full descriptive information about what variables have been measured. This template is mandatory.
Methods	The Methods template lists the methods used to collect the data and any additional information about the method. This template is mandatory.
Sites	The Sites template provides information giving the spatial location at which data values have been collected. This template is mandatory.
Sources	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.
Samples	The Samples template gives information about physical samples analyzed in a laboratory. This template is optional.
LabMethods	The LabMethods template contains descriptions of the laboratory methods used to analyze physical samples for specific constituents. This template is optional.
Quality Control	The QualityControlLevels template contains the quality control levels that are used for versioning data within the database. This template is mandatory.
DataValues	The DataValues template contains the actual data values and keys to metadata templates. This template is mandatory.
Categories	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.
DerivedFrom	The DerivedFrom template contains the linkage between derived data values and the data values that they were derived from. This template is optional.
GroupDescriptions	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.
Groups	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.
Qualifiers	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