A User's Guide to Registering and Maintaining Data Services in HIS Central 2.0

Prepared by Jon Pollak, CUAHSI Water Data Center User Support Specialist
September 2014

DISCLAIMERS

The HIS Central application and accompanying documentation are supported by the National Science Foundation under Grant No. 1248152. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Each user uses the WDC at her/his own risk. Data accessed through the WDC are provided "as is". CUAHSI does not verify the accuracy of data submitted by users. CUAHSI makes no representations or warranties, express or implied, with respect to the WDC and specifically disclaims all warranties to the fullest extent permitted by law (including, but not limited to, merchantability or fitness for a particular purpose). CUAHSI will not accept liability for damages of any kind that result from using the WDC.

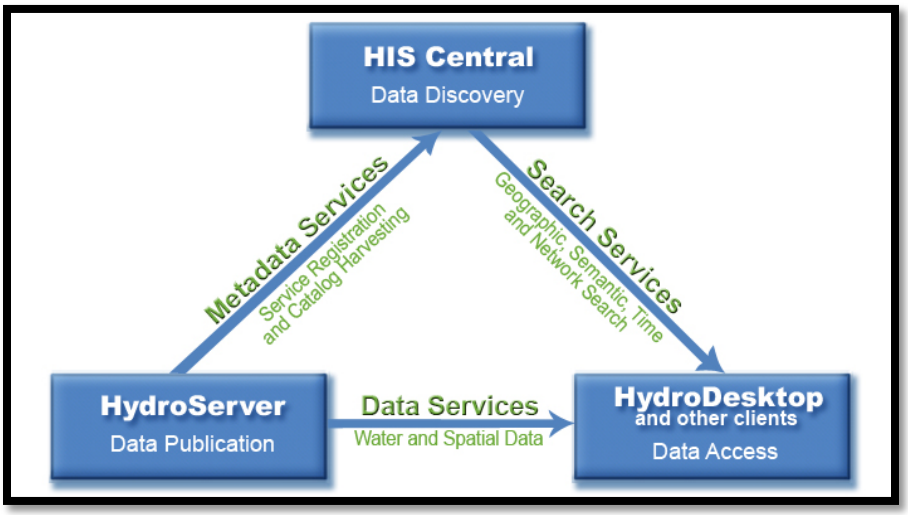
TECHNICAL SUPPORT

Ongoing technical support for HIS Central is provided by the CUAHSI Water Data Center. Contact CUAHSI staff by emailing help@cuahsi.org or calling 339-221-5400.

CONTENTS

Introduction to the Hydrologic Information System

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 works like common search services such as Google. HIS Central, the primary subject of this document, is a metadata data catalog (like the catalogs developed by Google). This catalog holds information about data sources (like Google holds information about websites) and one discovers and browses information using a client in the same way that one uses a web browser to browse websites.



The CUAHSI Hydrologic Information System (HIS) is designed for publishing time series data, that is observations collected over time at a fixed point, such as a record of stream stage or a series of water-quality observations at a station. These data are stored in a database called the Observations Data Model (ODM) which has defined a standard metadata profile for water data.

Within ODM, a time series is defined as collection of data having the same value for the following five attributes:

1. Site (e.g., USGS Site 08158000, Colorado River at Austin, TX)
2. Property measured (e.g., Alkalinity)
3. Method (e.g., Gran Titration)
4. Source (the collecting agency, e.g., University of Texas)
5. Quality Control Level (e.g., level of review of the data, e.g., “raw”, or “edited”)

These data sources are made available in the HIS through the registration of a web service URL on HIS Central. Registration, described later in this document, enables HIS Central to catalog the information contained in a data source makes the data source discoverable through search services using clients like HydroDesktop[1].

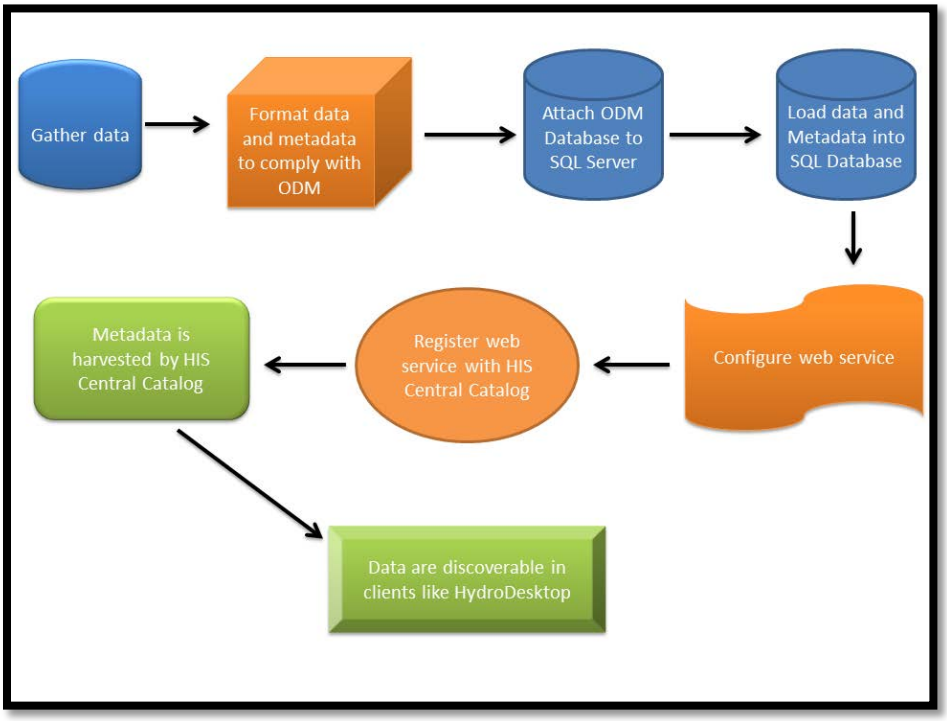
[1] http://www.hydrodesktop.org

Overview of HIS Central

HIS Central is the web application which provides an interface for adding and managing registered water data services and the HIS central metadata catalog. The central metadata catalog is designed to maintain observation series information, including site information, variable information, period of record, as well as project metadata – for all registered data sources of hydrologic observations. These two components are in the center of the CUAHSI HIS architecture.

There are several user roles associated with HIS Central. Data managers, who have gone through the task of creating a WaterOneFlow data service for their observational data, use it to register the service to the central registry, to make it possible for HIS users to discover and access the data, and integrate it with information from other similarly registered data sources. The HIS Central Administrators are responsible for maintaining hydrologic metadata from sources other than those compatible with WaterOneFlow services such as U.S. federal agency repositories. In addition, this role is responsible for periodically updating the metadata catalog and managing all aspects of the HIS Central application. The metadata catalog also maintains a hydrologic ontology so that well established concepts can be used to search across multiple water data sources, which may be using different terminology for their respective variables. An ontology curator is responsible for maintaining and evolving the ontology, and managing associations between ontology concepts and variables. Once data are published and registered in HIS Central, they are accessed by metadata catalog users, via online and desktop clients. In addition to the concept-based search enabled by the ontology mappings, the metadata catalog maintains spatial and temporal indexes assisting in spatio-temporal searches across multiple observation series.

Workflow for Data Publication



Generally, there are seven steps that comprise the workflow for publishing data in the HIS, which are displayed in the flow chart above. The role HIS Central plays is in making existing data available through web services. Before this occurs, however, a researcher must gather or otherwise create data and format the data to comply with the HIS' information model, The Observations Data Model[2] (ODM). Typically, the ODM is manifested as an SQL Relational Database. Next, the WaterOneFlow Web Service is configured and connected to the database. The web service creates a URL endpoint that can be registered with HIS Central. This permits a web crawler from HIS Central to index the contents of the database and enables discovery and download of the database's contents.

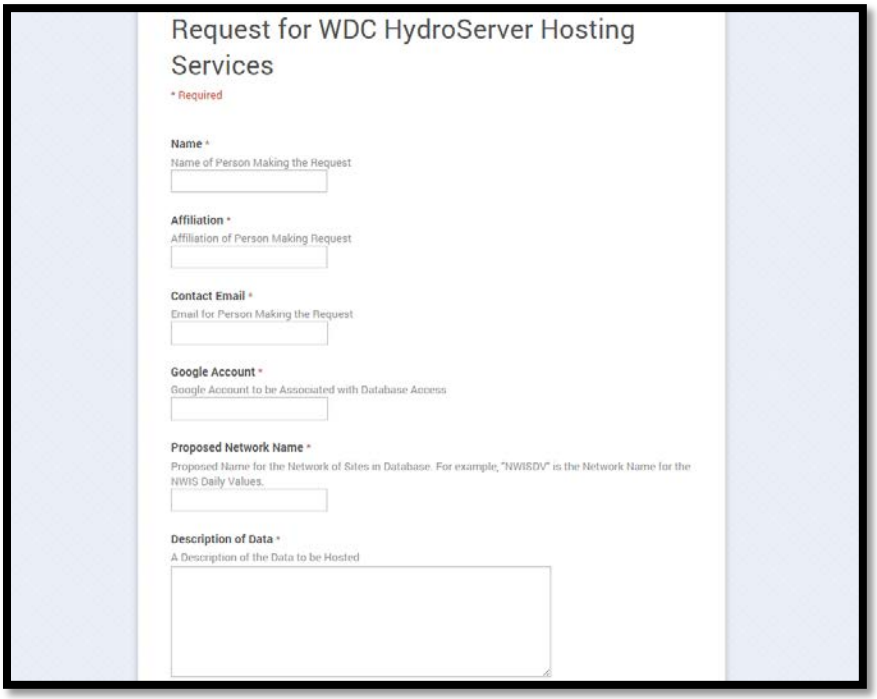
The WDC provides hosting services that simplify this workflow for many researchers. WDC staff will setup the database and web service for a respective researcher or group, but it remains the researcher or group's responsibility to format the data properly as well as to complete the registration process with HIS Central. To request hosting services, visit http://histools.cuahsi.org.

[2] Horsburgh, J. S., D. G. Tarboton, D. R. Maidment, and I. Zaslavsky (2008), A relational model for environmental and water resources data, Water Resour. Res., 44, W05406, doi:10.1029/2007WR006392.

Adding a New Data Service

WDC Hosting Services

If you are hosting your own database and web service externally from the CUAHSI WDC, please skip to the next section, *Registering a New Service*. This section briefly describes the process of requesting hosting services from the WDC, which includes the information that the WDC staff will need, and the information the WDC staff gives to the data publisher that is needed to complete a registration.



To request hosting services from the WDC for time series data to be registered with HIS Central, visit http://histools.cuahsi.org. You will be prompted to log in with a Google Account. If your Google Account does not have a database associated with it then you will be brought to a request form. Provide the required information and CUAHSI staff will contact you once your information has been processed.

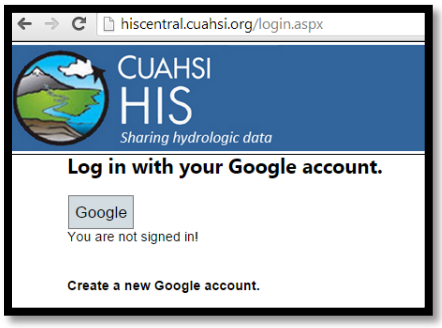
If you need assistance, or do not have a Google Account, contact CUAHSI staff by emailing help@cuahsi.org.

Registering a New Service

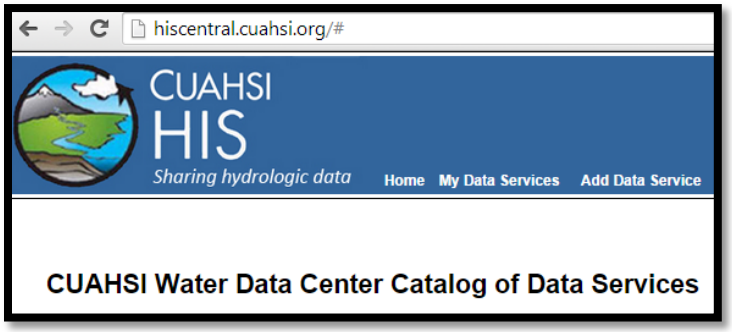
The following list contains the steps necessary for registering a new data source with HIS Central.

Below we present an outline of a step-by-step process that data managers are expected to follow when registering their data in the HIS Central.

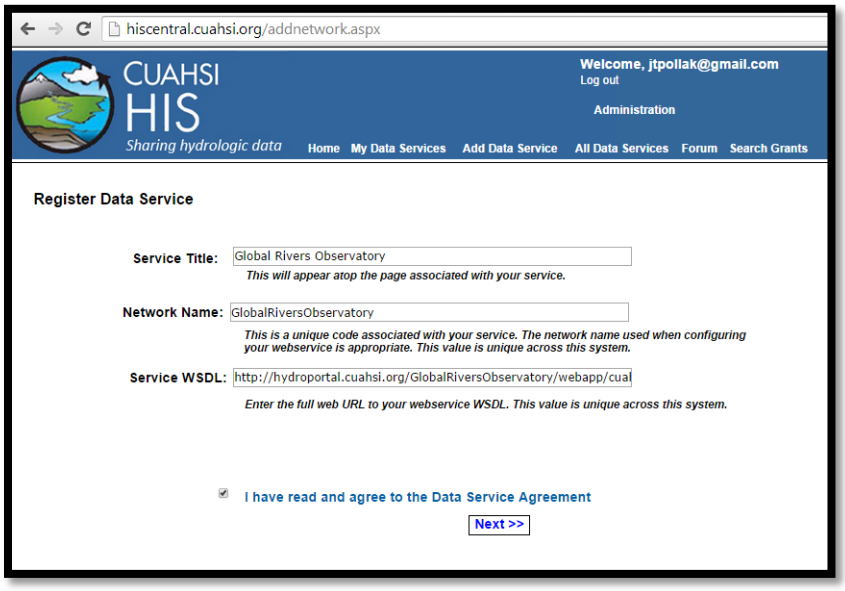
1) Login to HIS Central using a Google Account at http://hiscentral.cuahsi.org



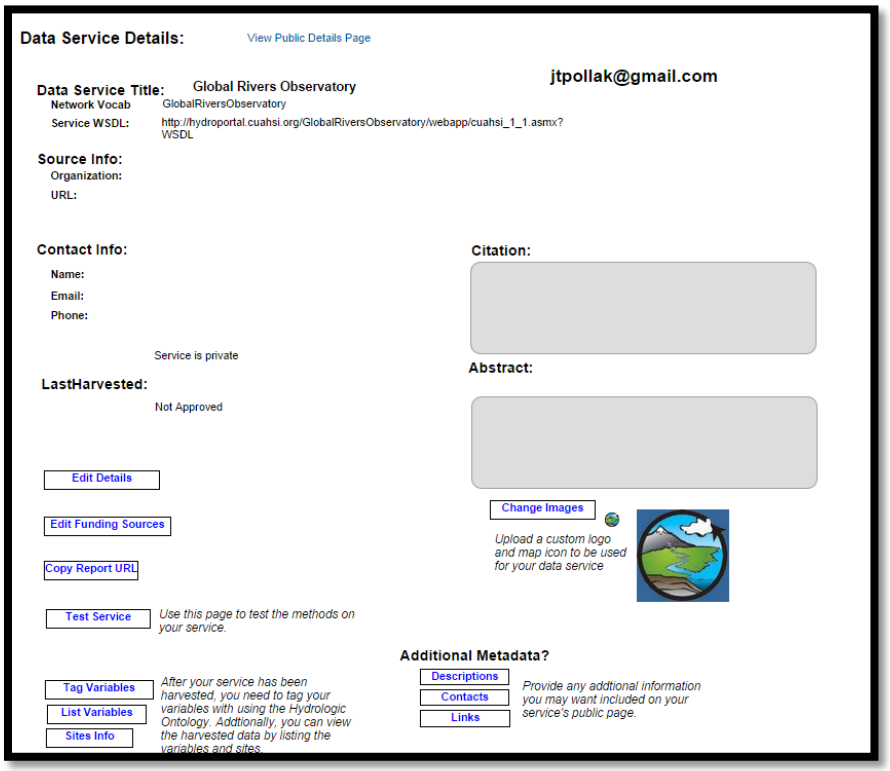
2) Click Add Data Service

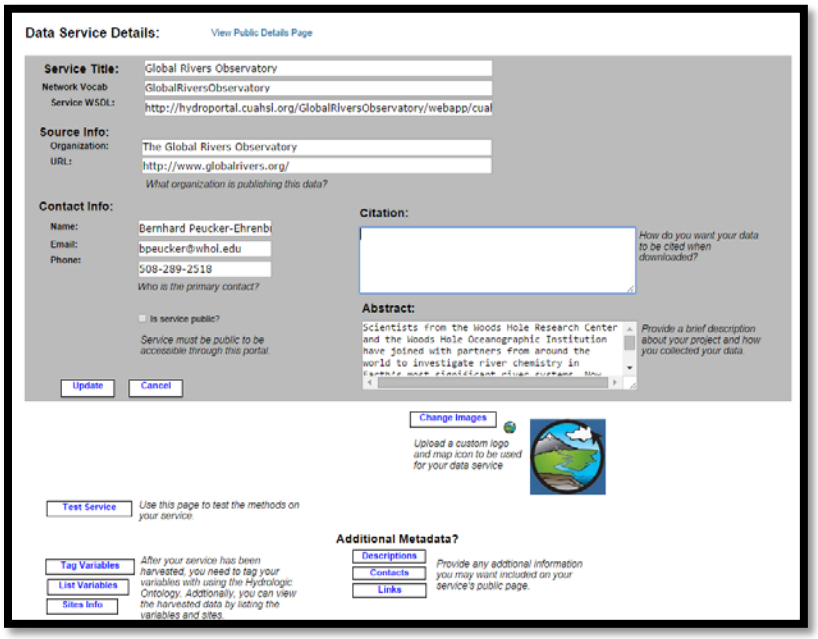


3) Enter Data Service Information

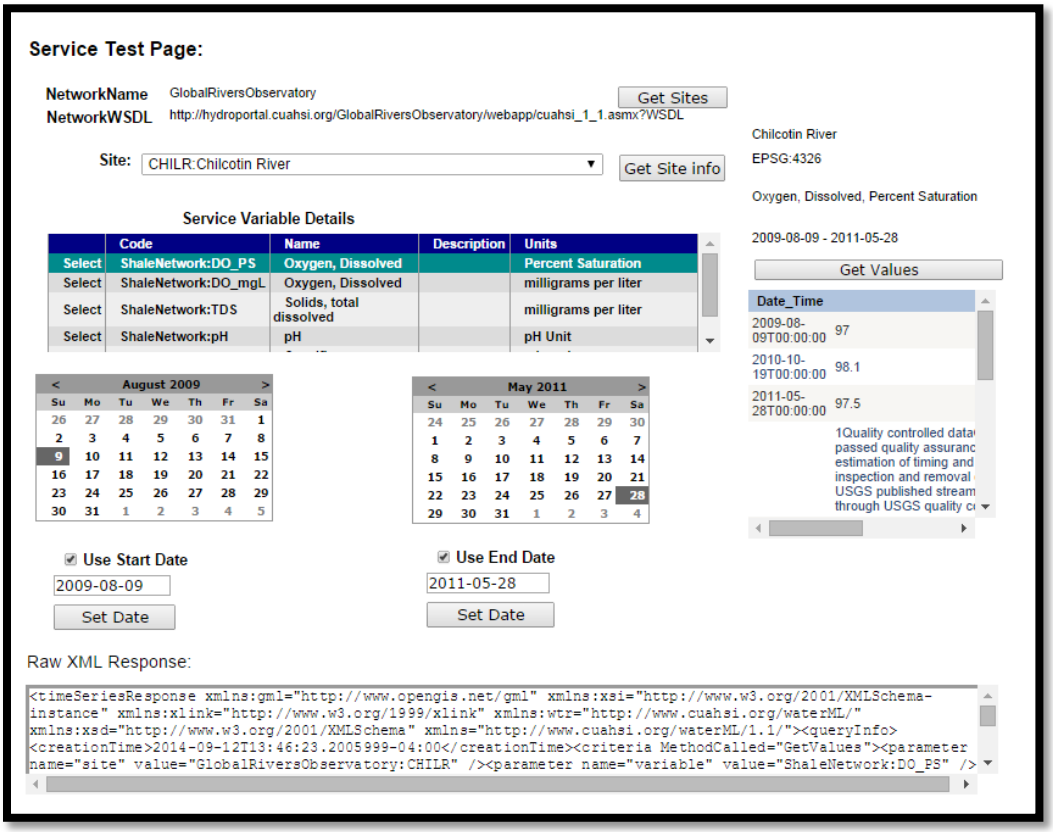


4) Enter Data Service Metadata

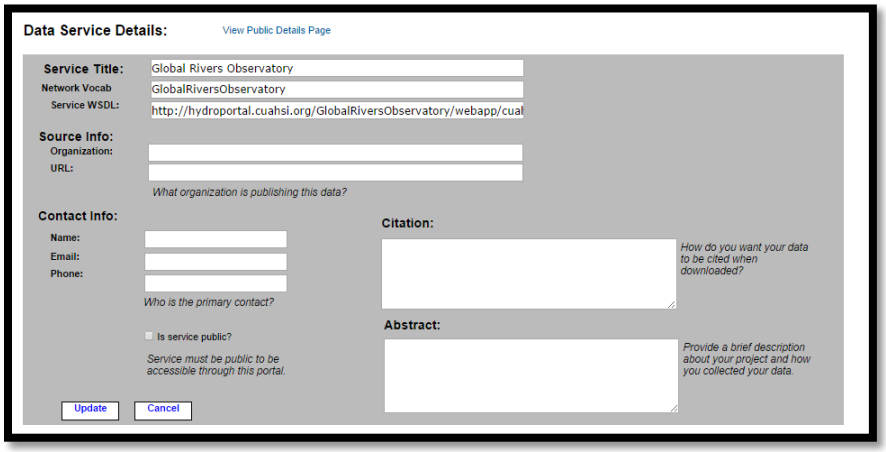




5) Test the data service using the test page to make sure data values and metadata appears as expected. There is additional information for testing services in the next section of this document, *Managing your Data Service(s)*.

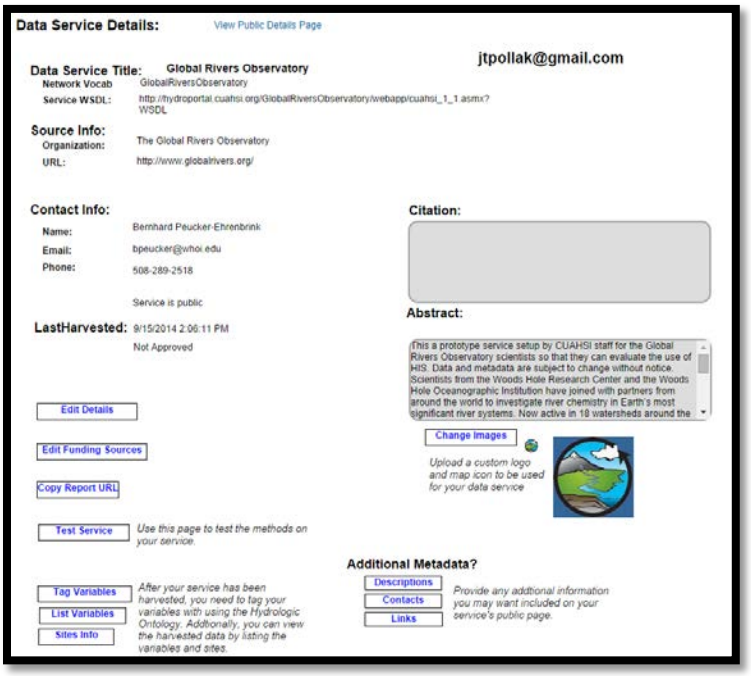


5) Change the status of your service to public in order to make data retrievable in search results.



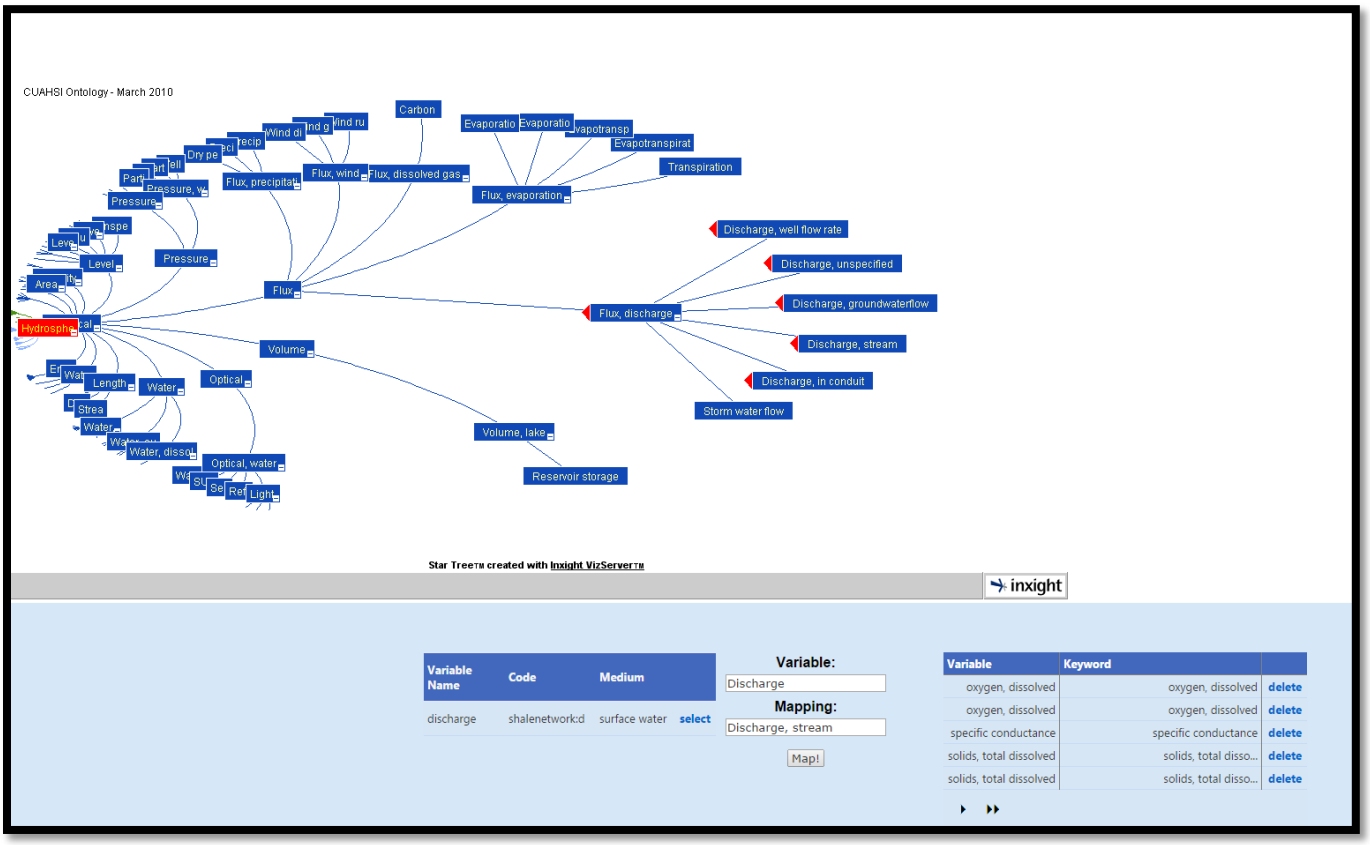
6) A metadata harvester operates every Saturday, which downloads metadata from the data service in order to enable search and discovery of data. A data service will not appear in search results until it has been harvested. To request a harvest before the weekend, email help@cuahsi.org.

7) Verify Tagging of Variables to Ontological Concepts. HydroTagger is the name of the tool used to tag the variables in an ODM database to the concepts of HIS Central's ontology and is accessible from each data service's service details page by clicking the Tag Variables link, which can be seen in the lower left of the screenshot below. **Note:** If a data publisher conforms to using CUAHSI's Controlled Vocabulary[3] then the majority of variables will be mapped automatically. It is the data publisher's responsibility to review these mappings to ensure that they are acceptable to the data publisher.



[3] http://his.cuahsi.org/mastercvreg/edit_cv11.aspx?tbl=VariableNameCV&id=821577965

To map a variable, select it in the table in the lower left of the page. Next, pan the hierarchy or terms or search for a term in the *Search* box. Search results will be highlighted with red triangles, which can be seen in the screenshot below. Select a term to map the variable to by double clicking on the respective term in the hierarchy and clicking *Map!* To delete a mapping, click the *delete* button next to the applicable Variable and Keyword.

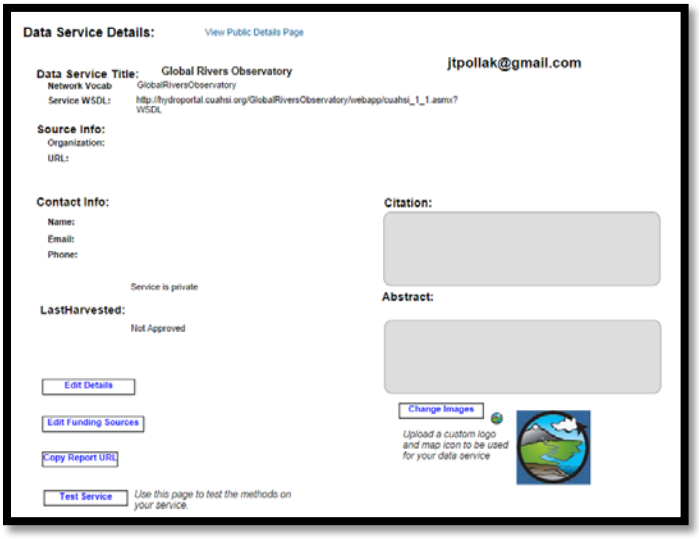


8) Once your variables are tagged, your data are available to be discovered and downloaded in data access clients like HydroDesktop.

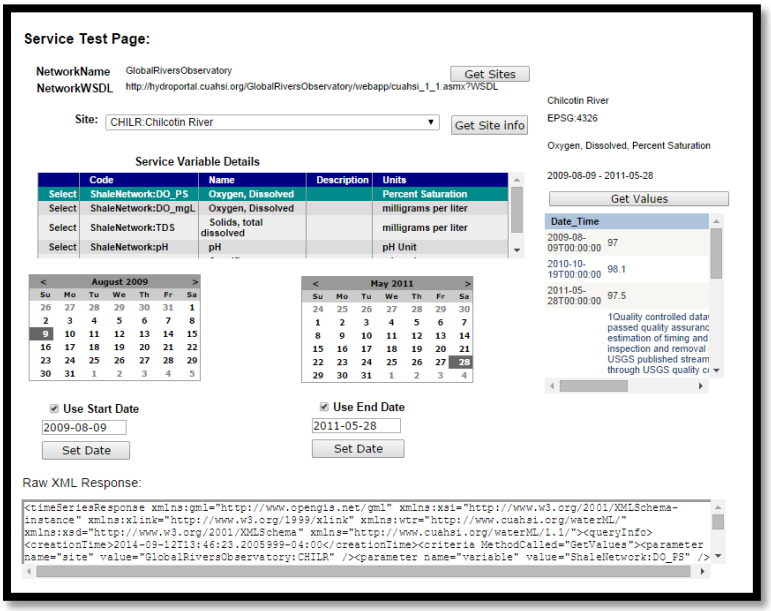
MANAGING YOUR DATA SERVICE(S)

TESTING A DATA SERVICE

TESTING A SERVICE ON HIS CENTRAL



To test a data service at any time, go to the details page for the specific service and click Test Service. The resulting page will contain an interface for querying for the sites in your service, variables at a specific site, and finally data values for a specific variable at a specific site filtered by date. The page also displays the raw XML response for the query at the bottom of the page.



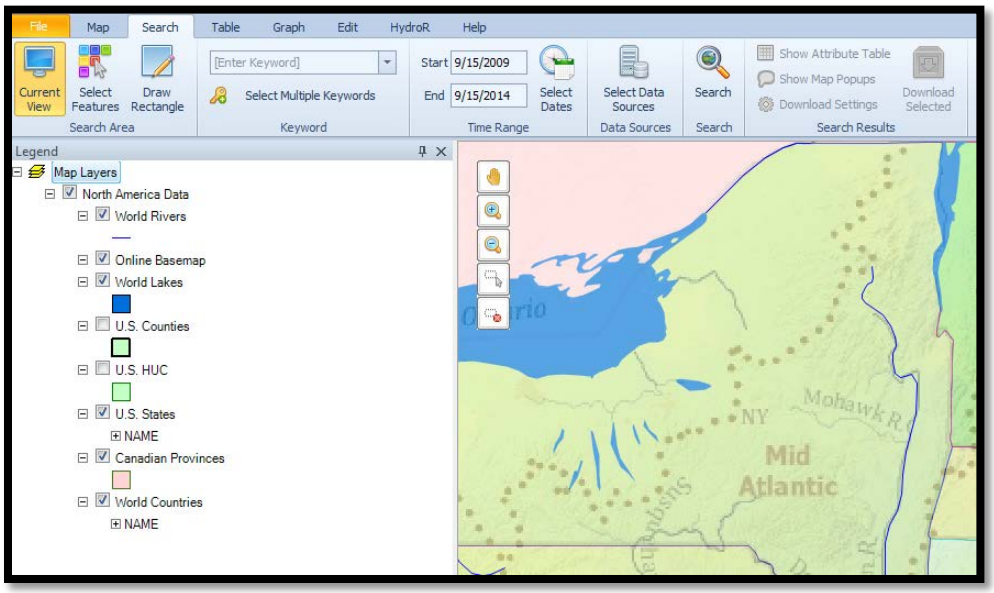
This test feature is available whether or not a specific data source is public. It is especially useful for data managers who are managing their own server because it will verify that the server is responding correctly to WaterOneFlow web service calls. If the server is not responding correctly, data will not be returned with this interface or data access clients like HydroDesktop.

TESTING A SERVICE IN HYDRODESKTOP

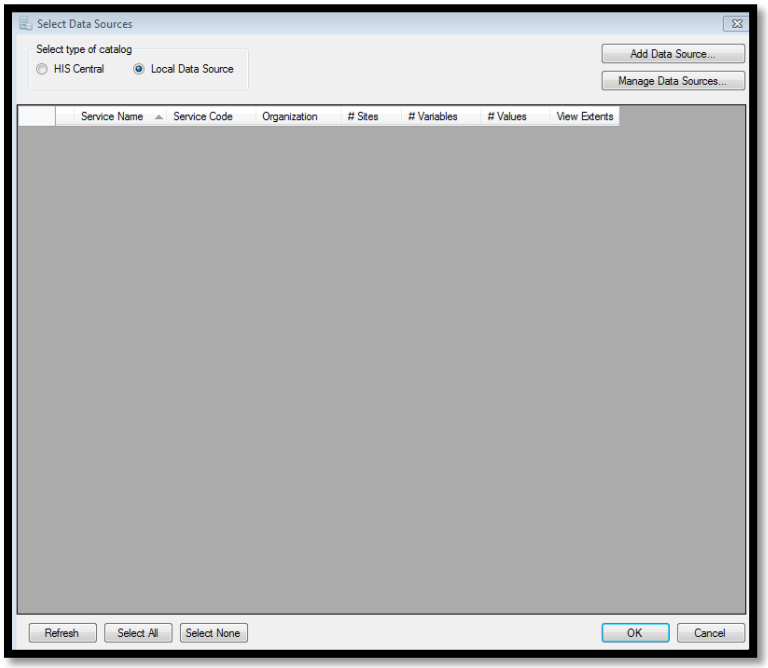
Another option for testing a data service, in addition to the process described above in HIS Central, is to test a service in HydroDesktop. This can be completed once the service has a valid WSDL endpoint and can be either prior to or after the data service has been registered with HIS Central.

HydroDesktop is a free, open source GIS software application for Windows that can interact with the other pieces of the HIS. As of the writing of this document the latest version in HydroDesktop 1.6. The workflow described below, and screenshots, are from this version.

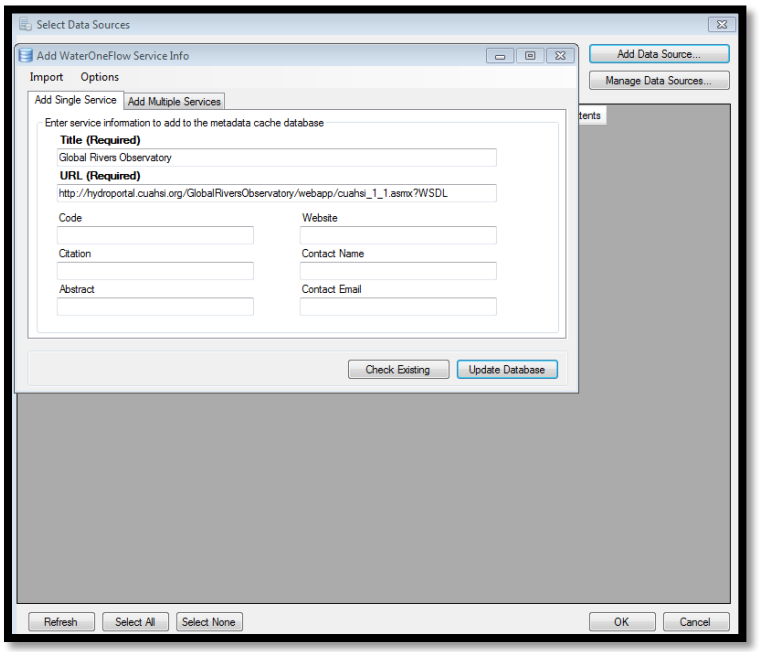
The first step for testing a service in HydroDesktop is identifying the location of the service. Under the Search tab in HydroDesktop, click the Select Data Sources button.



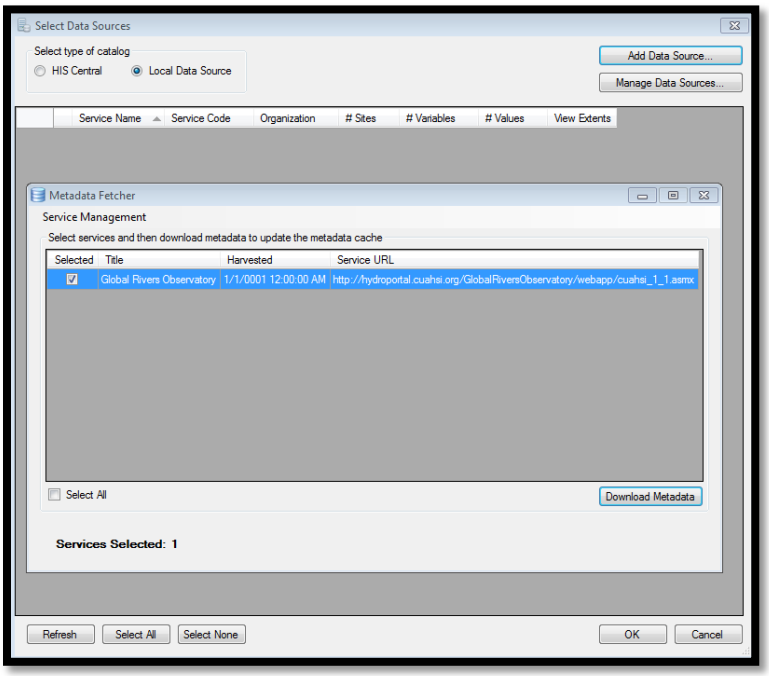
In the Select Data Sources window, choose the Local Data Source radio button. The Local Data Source screen enables the direct input of a WSDL for a data service.



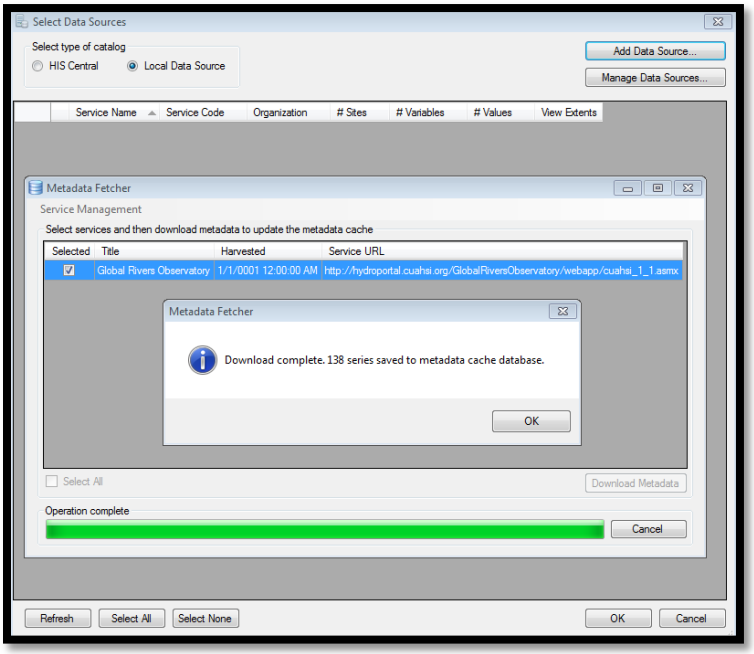
The Add Data Source button will bring up a new screen where a Title and URL (for the WSDL) can be entered.



The last step to bring this data source into HydroDesktop is to download the metadata for the service to HydroDesktop's local storage. Click Update Database then choose to download the metadata from the service in the Metadata Fetcher window.



If this download is successful, then HIS Central's harvester will be able to harvest the metadata from the service during the registration process.



At this point, the service is discoverable locally using the search tools in HydroDesktop. The data are not publicly available until the HIS Central registration process is complete, which is described in the preceding section of this document.

Viewing Metrics for a Data Service

Metrics for data services can be accessed from the My Data Services page by clicking Usage Report (located on the left).

The resulting page will display a graph of usage statistics and contains links to download the metrics as a .csv file.

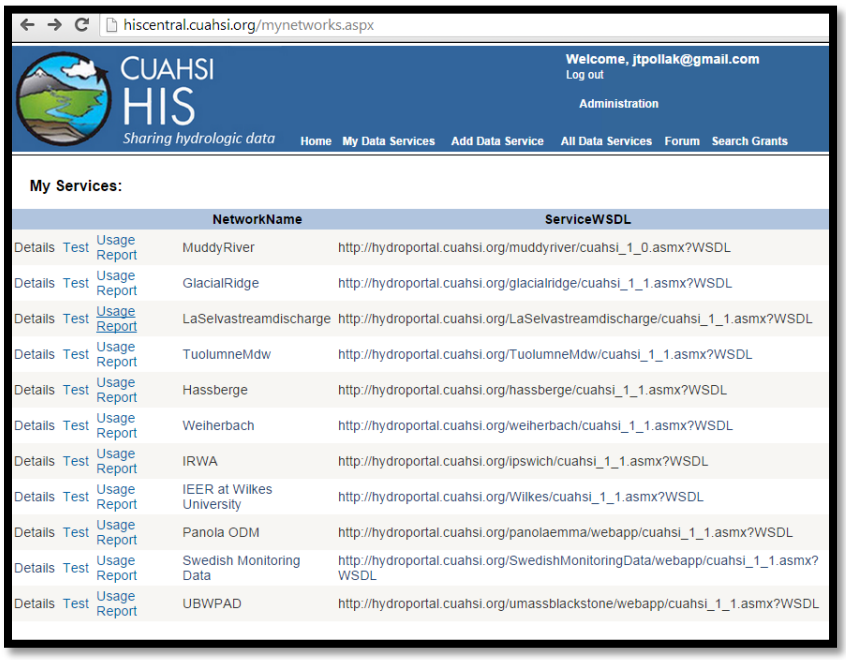
If there are no metrics being displayed, ensure that the Network Name used during service registration matches the Network Name defined in the WaterOneFlow Web Service Application. Need assistance accessing metrics? Let us know: help@cuahsi.org.

Adding and Editing Funding Sources

A new feature in HIS Central 2.0 enables a data publisher to tag funding sources to Organizations from the Sources table in ODM. For example, if an ODM database has two organizations contributing data and one has been funded

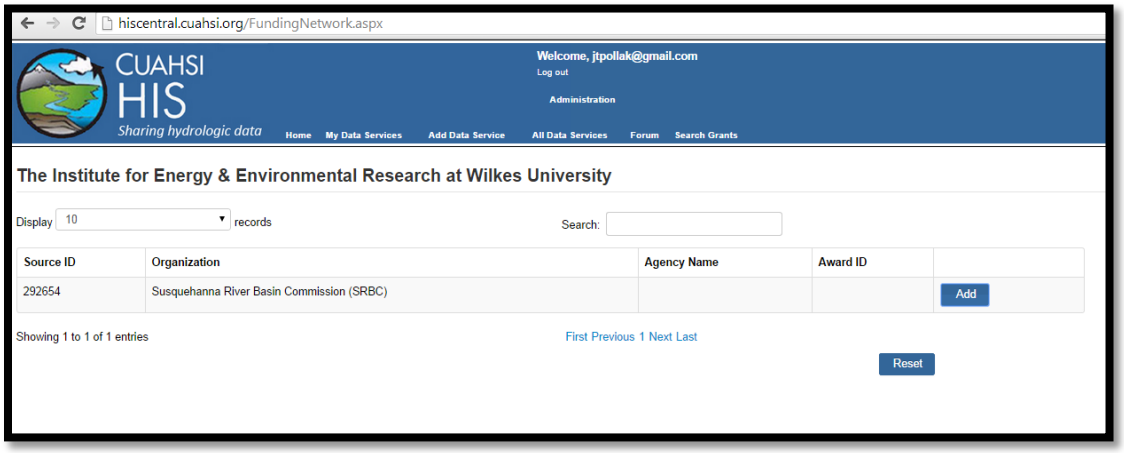
by the National Science Foundation while one was funded by the U.S. Environmental Protection Agency, both funding sources can be acknowledged with the applicable grant number and supporting hyperlinks. To add a funding source, follow the steps outlined below.

1. From the My Data Services Page, click the Details link in the column farthest to the left in order to access the service details page for the data service you wish to add/edit funding information for.

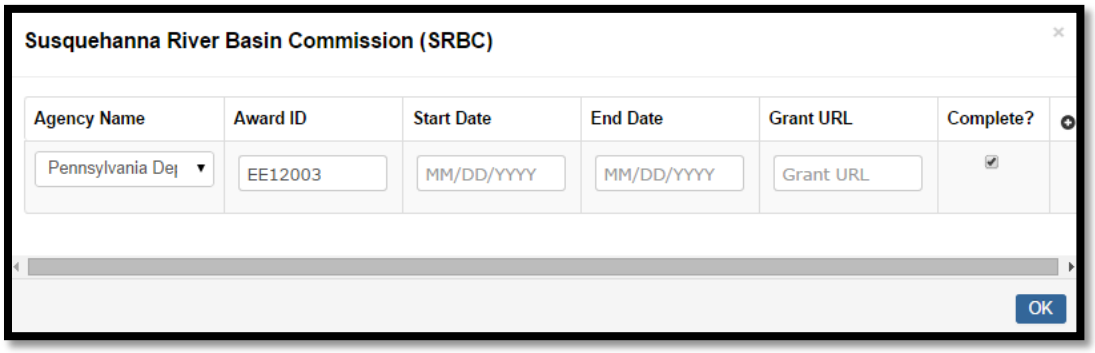


2. From the Service Details Page, click Edit Funding Sources.

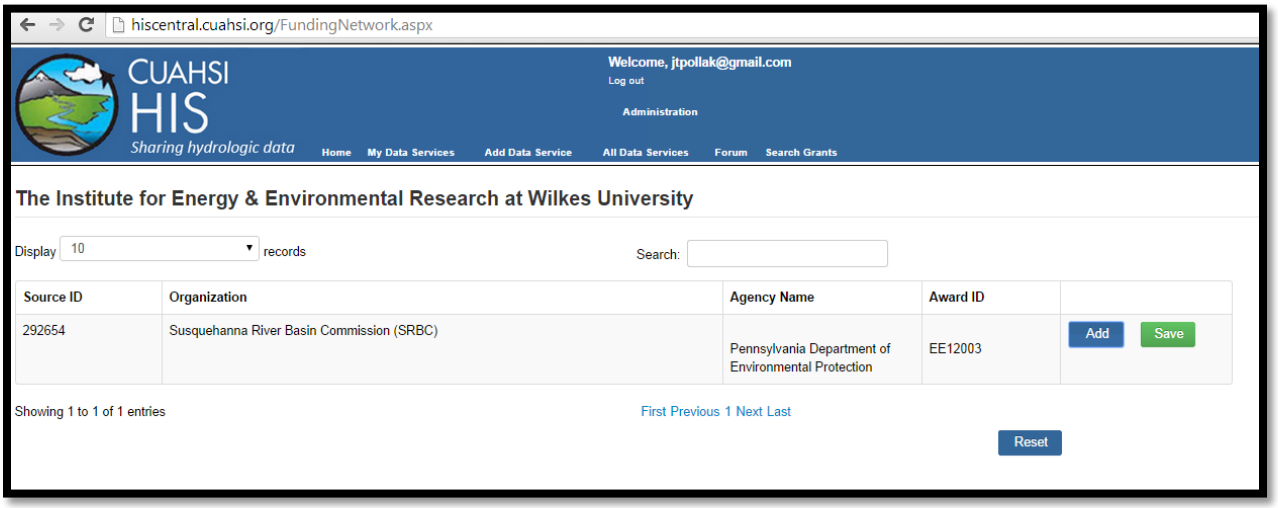
3. The resulting page will display all Organizations included in the Sources table of the ODM database for the service. Click Add/Edit on the right side of the table to add or change funding information for each organization.



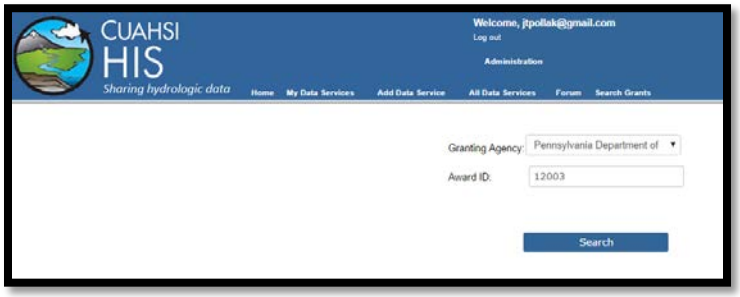
4. Select an Agency Name, enter the AwardID, Start Date (optional), End Date (Optional), Grant URL (optional) and check the box to the far right if the grant is complete.



5. Click OK in the table, and then click Save to commit your edits to the database.

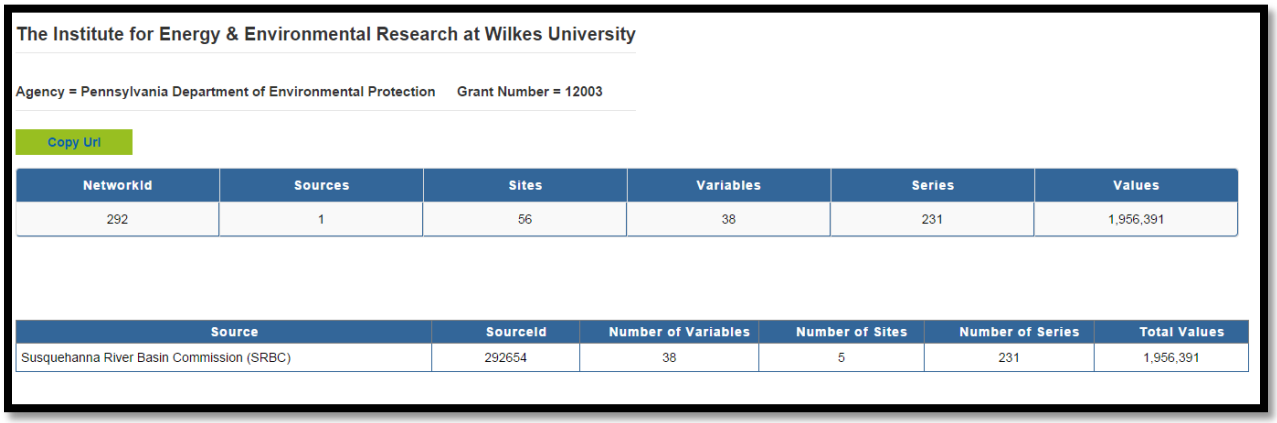


6. Once the funding information has been saved it is searchable via the Search Grants page, the link for which is located on the header of HIS Central. To search for a data service based upon its granting agency and grant number, go to the Search Grants and enter the Granting Agency and Award ID.



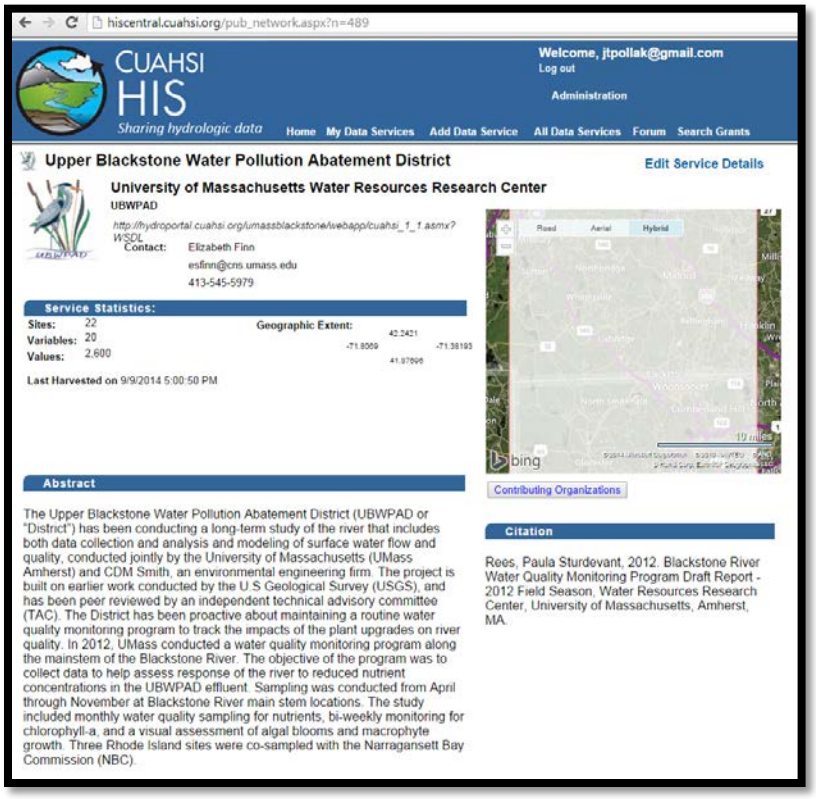
7. A successful search will yield a report webpage with simple statistics related to the grant

8. Click Report to access a statistical report about the data associated with the grant. This page has a unique URL that is meant to be included in reports such as those required by the National Science Foundation.



Browsing the Catalog and Downloading Data

The HIS Central webpage[4] contains a list of all public data services registered with the catalog. Additionally, each service has a service details page that contains information about the data service including statistics, geographic extents, as well as a description or summary of the data source.



HydroDesktop

The most common tool used for data discovery and download for data registered in HIS Central is HydroDesktop. To download the most recent version of HydroDesktop, or for more information, visit www.hydrodesktop.org.

[4] http://hiscentral.cuahsi.org