An Overview of the Hydrologic Information System and CUAHSI Water Data Center

Jon Pollak

Universidad de Los Andes

4 December 2013

Purpose

- Background on CUAHSI and the CUAHSI Water Data Center
- Overview of Hydrologic Information System
- Case Studies of Data Publication
- Near Future Plans for the Data Center

What is CUAHSI?

- Consortium of Universities for the Advancement of Hydrologic Science, Inc.
- Nonprofit primarily funded by the U.S. National Science Foundation to promote and advance hydrology science and education
- Founded in 2001
- Over 100 members
  - Primarily US universities, international affiliates, and private companies

National Science Foundation

- US Federal Agency that promotes scientific research
  - Provides resources for academic community through competitive grants
- Provides majority of funding for CUAHSI

What is the CUAHSI Water Data Center?

- Facility funded by the National Science Foundation
- Technology and data catalog developed through a multiyear, multi-university research project

CUAHSI WDC Activities

- **Data catalog:** creation, maintenance, services and curation
- **User and developer support**.
  - Assistance with publishing new data sources and/or software
- **Strategic partnerships** with academia and industry
- **Outreach and education**

Who am I?

- User Support Specialist for the Hydrologic Information System at the CUAHSI Water Data Center
  - Help researchers find data
  - Help researchers publish data
  - Aid in product development by working with
    - Software engineers
    - Users Committee

Hydrologic Information System Research Project

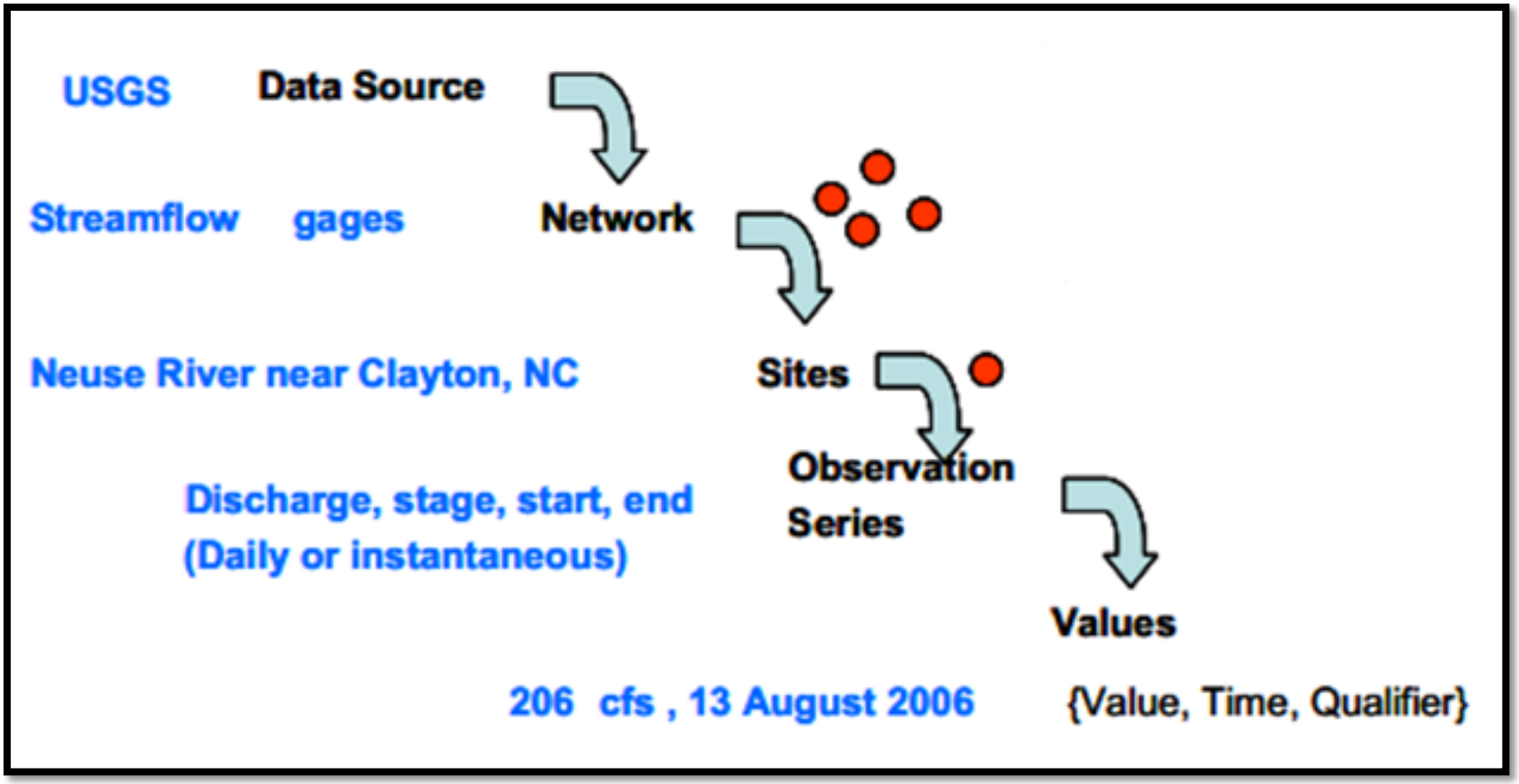
- Led by Professor David Maidment (University of Texas at Austin) with…
  - Utah State University
  - San Diego Supercomputer Center
  - Idaho State University
  - City College of New York
  - University of South Carolina
  - CUAHSI

CUAHSI HIS

Sharing hydrologic data

Hydrologic Information System Research Project

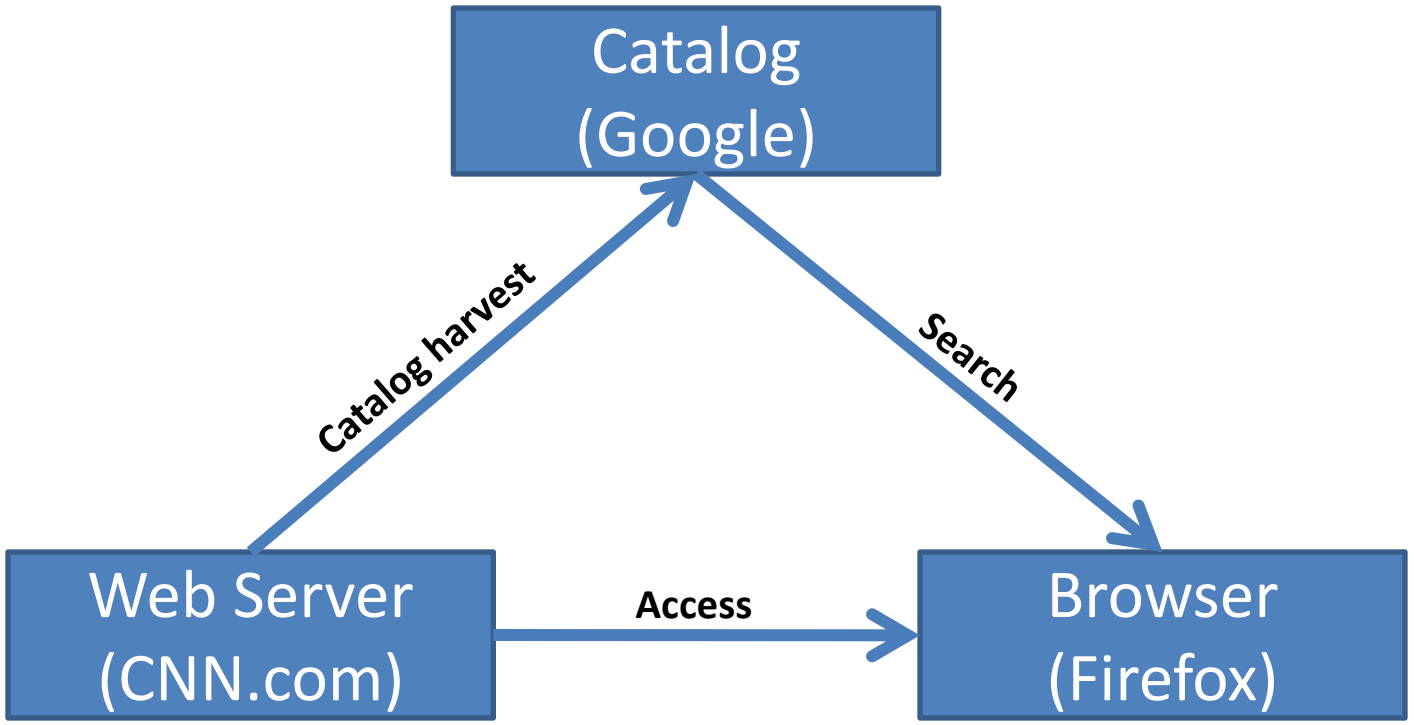
- A system for sharing time series water data
  - Observations at a single location over time



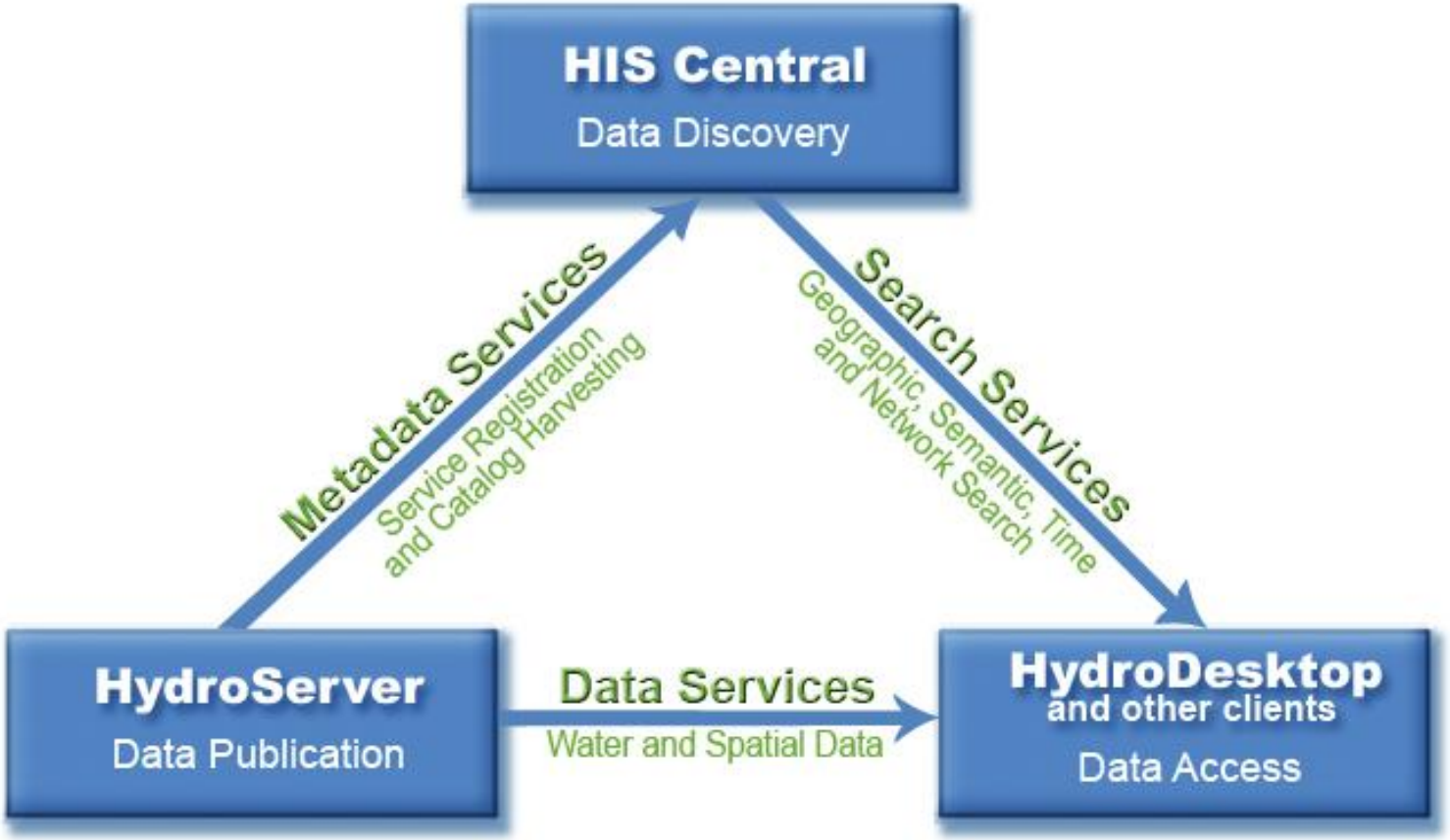
Some findings from the CUAHSI HIS Reseach Project

- **Use Service-Oriented-Architecture** (SOA) to provide data to discovery clients.
- **Define standards** for data publication (WaterML).
- **Write data discovery clients** that utilize the standards.
- **Adapt existing data sources** to conform to standards by writing translators between existing and standard data formats.
- **Define an ontology** of data concepts to enable cross-domain use and data discovery, including synonyms and concept relationships.
- **Include terms and synonyms** that describe data in a variety of scientific contexts.

Web Paradigm



Service Oriented Architecture



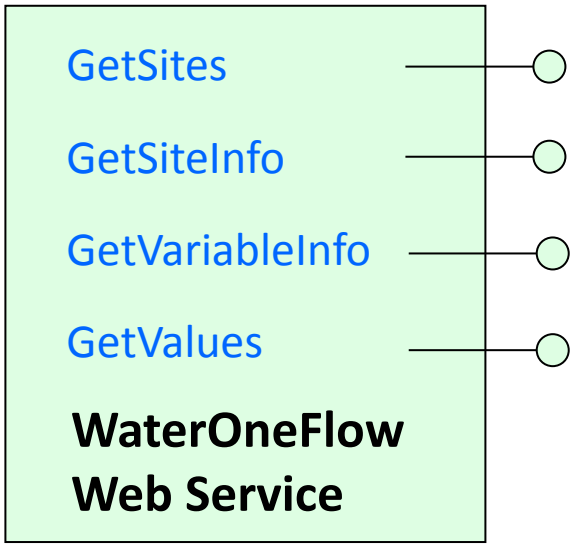
Web Services

- Software which is a method of communication between two computers over the internet
- WaterOneFlow Web Service is a suite of web services employed the HIS
- Data are returned in a type of Extensible Markup Language (XML) developed during the HIS Project called WaterML.

WaterML and WaterOneFlow

WaterML is an XML language for communicating water data
WaterOneFlow is a set of web services based on WaterML

- Set of query functions



- Returns data in WaterML

```
<timeSeries>
- <sourceInfo xsi:type="SiteInfoType">
    <siteName>Colorado Rv at Austin, TX</siteName>
    <siteCode network="NWIS" siteID="4619631">08158000
  - <geoLocation>
    - <geogLocation xsi:type="LatLonPointType" srs="EPSG
        <latitude>30.24465429</latitude>
        <longitude>-97.694448</longitude>
      </geogLocation>
    </geoLocation>
  </sourceInfo>
- <variable>
    <variableCode vocabulary="NWIS" default="true" variable
    <variableName>Discharge, cubic feet per second</vari
    <units unitsAbbreviation="cfs" unitsCode="35">cubic fee
  </variable>
- <values count="2545">
    <value dateTime="2006-12-31T00:00:00">129</value>
    <value dateTime="2006-12-31T00:15:00">129</value>
    <value dateTime="2006-12-31T00:30:00">129</value>
    <value dateTime="2006-12-31T00:45:00">129</value>
    <value dateTime="2006-12-31T01:00:00">124</value>
    <value dateTime="2006-12-31T01:15:00">129</value>
    <value dateTime="2006-12-31T01:30:00">124</value>
    <value dateTime="2006-12-31T01:45:00">124</value>
```

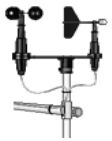
HydroServer

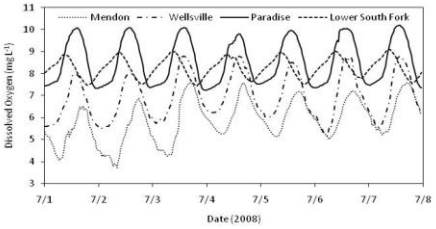
- Core components are:
  - Database (most commonly Microsoft SQL)
  - WaterOneFlow Web Service
- Additional software includes:
  - Data Management software
  - Website application
  - Map server application

HydroServer – Data Publication

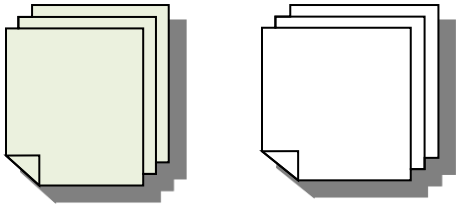
Point Observations Data

Ongoing Data Collection



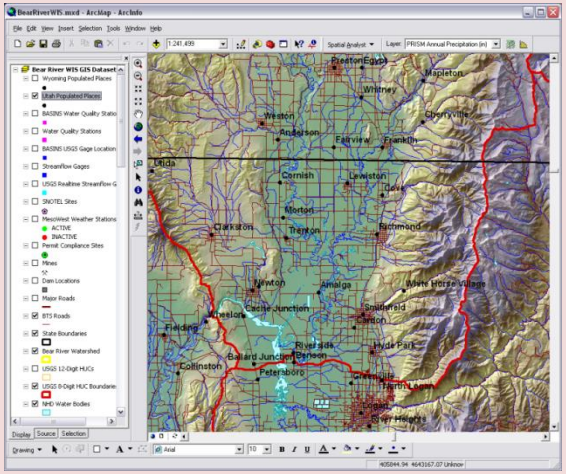


Historical Data Files



ODM Database

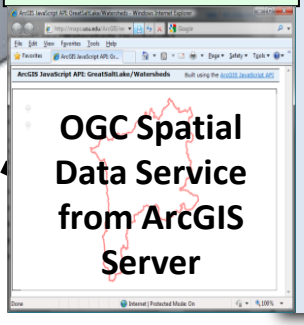
GIS Data



GetSites
GetSiteInfo
GetVariableInfo
GetValues

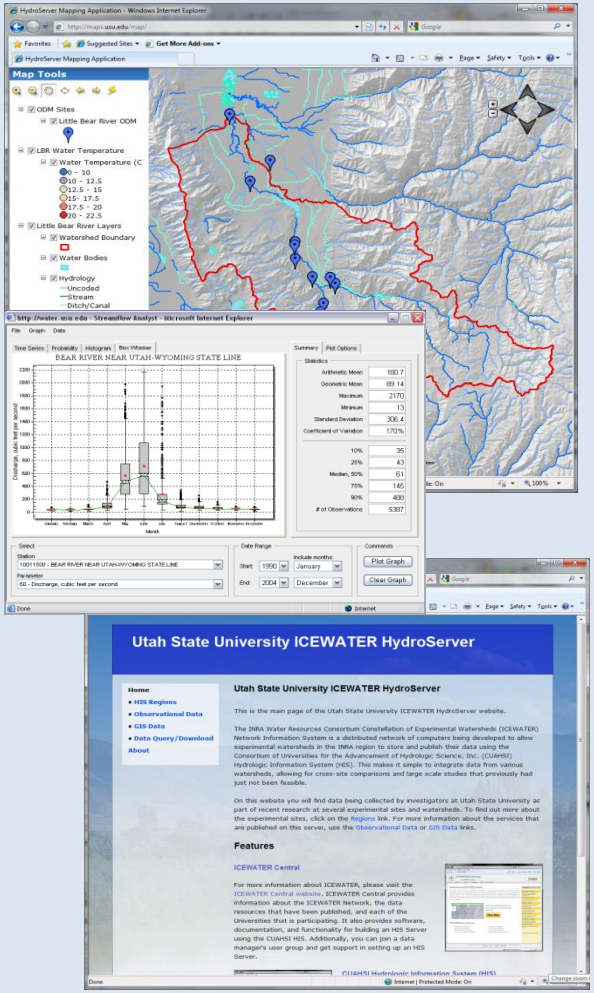
WaterML

WaterOneFlow Web Service



OGC Spatial Data Service from ArcGIS Server

Internet Applications

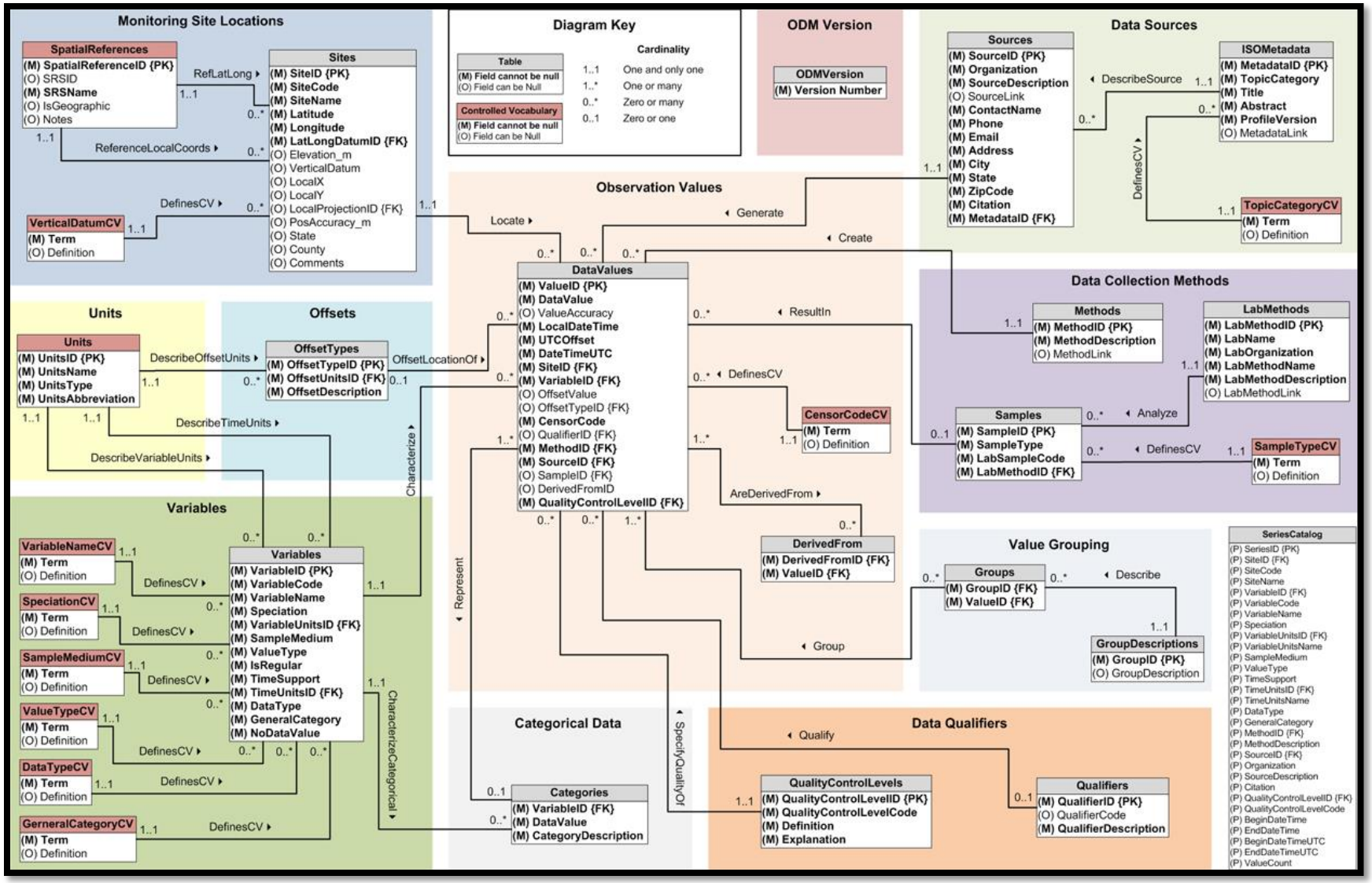


Data presentation, visualization, and analysis through Internet enabled applications

The Observations Data Model

- Information model which the system is based upon
- Purpose is to enable data storage that optimizes data retrieval for integrated analysis of information collected by multiple investigators
- For detailed information:
  - Horsburgh, J. S., D. G. Tarboton, D. R. Maidment and I. Zaslavsky, (2008), *A Relational Model for Environmental and Water Resources Data*, Water Resources Research, 44 (5).

Observations Data Model

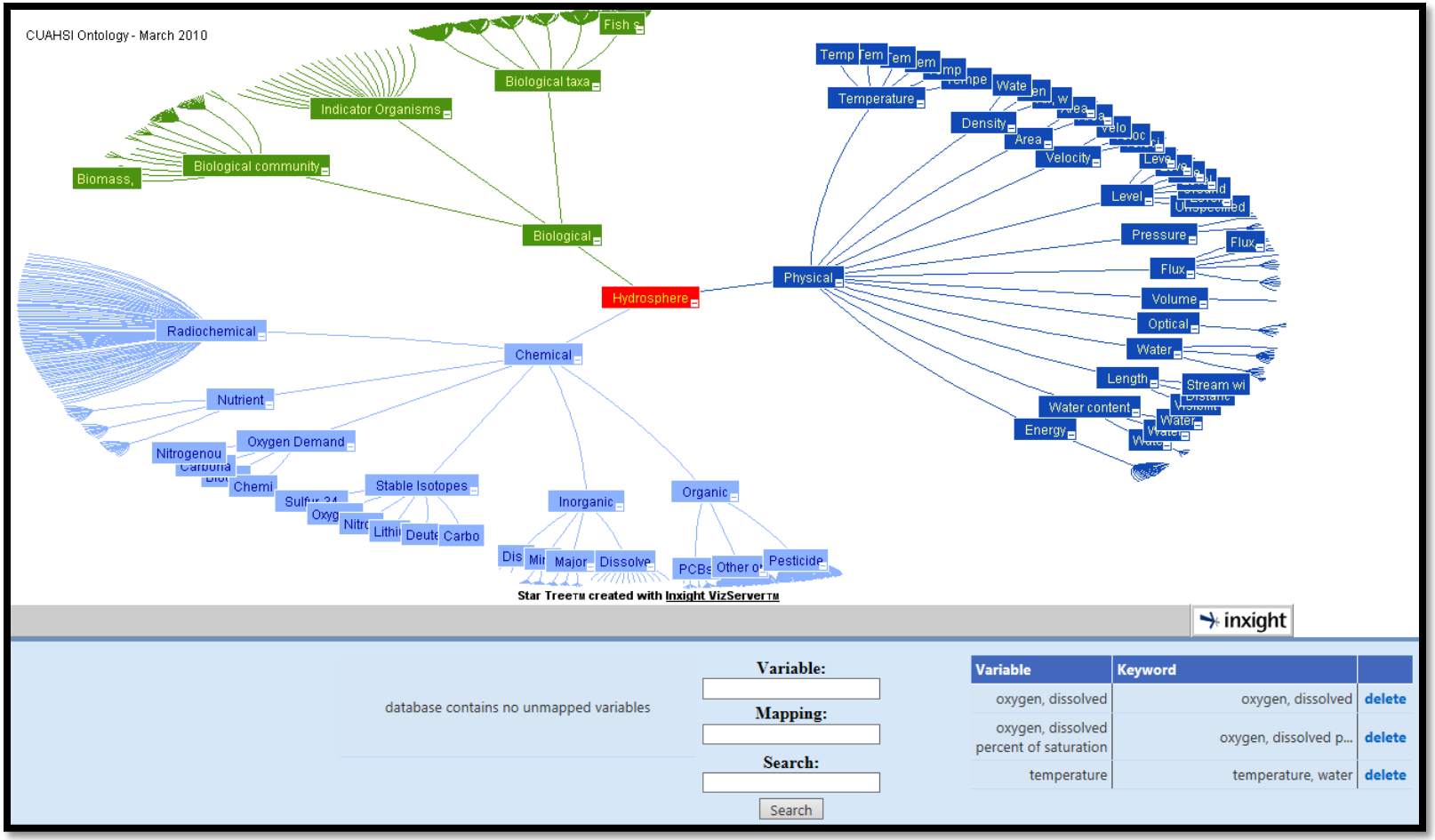


Most common form is Microsoft SQL Database, but MySQL also used

HydroCatalog

- Catalog that enables search and discovery
- Stores metadata about each data source
- Users register web service with the catalog, catalog uses a program similar to a web crawler to harvest metadata
  - Publisher also manually enters metadata about the data source

HydroCatalog



- Ontology
  - Users “tag” variable (i.e. temperature) to terms in the ontology, which enables keyword search

HydroCatalog

hiscentral.cuahsi.org/pub_network.aspx?n=52

CUAHSI HIS
Sharing hydrologic data

Login Register

Home All Data Services

Little Bear River Experimental Watershed, Northern Utah, USA

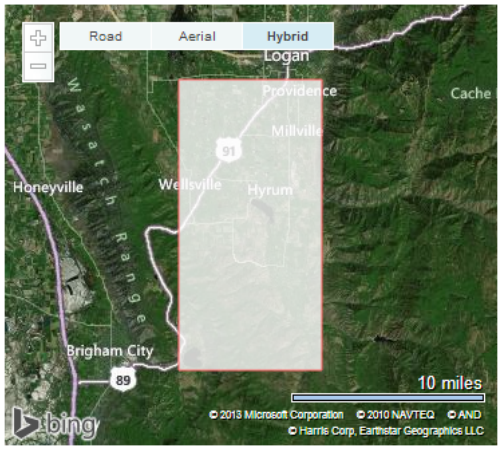
Utah Water Research Laboratory, Utah State University

UtahState UNIVERSITY

LittleBearRiver

WaterML Service:
http://icewater.usu.edu/littlebearriver/cuahsi_1_1.asmx?WSDL
WFS Service:
http://hiscentral.cuahsi.org/WFS/52/cuahsi.wfs?request=getCapabilities

Contact: Jeff Horsburgh
jeff.horsburgh@usu.edu
435-797-2946



Service Statistics:

Sites: 16
Variables: 61
Values: 23288662

Geographic Extent:
41.71847
-111.9464 -111.7993
41.49541

Last Harvested on 11/30/2013 12:15:50 PM
(updated weekly, assumed current)

Abstract

Utah State University is conducting continuous monitoring within the Little Bear River watershed of northern Utah, USA to investiage the use of surrogate measures such as turbidity in creating high frequency load estimates for constituents that cannot be measured continuously.

Keywords:

Discharge, Water Quality, Pollutant Loads, Continuous Data, Surrogate Measures, Oxygen Dynamics, Hydrochemical Response

Citation

Horsburgh, J. S., D. K. Stevens, D. G. Tarboton, N. O. Mesner, A. Spackman Jones, and S. Gurrero (2009) Monitoring data collected within the Little Bear River Experimental Watershed, Utah, USA, Utah State University.

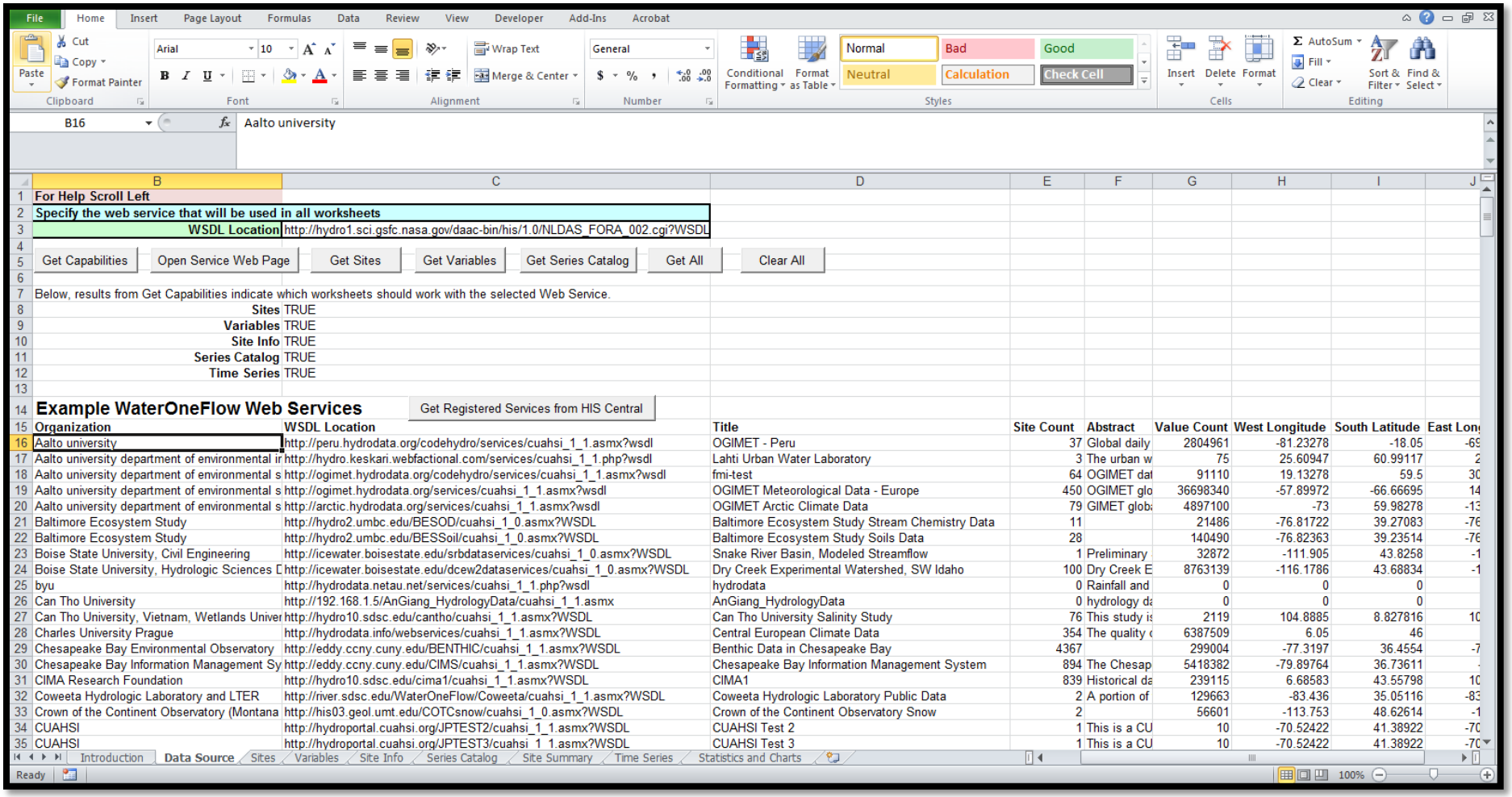
Public webpage for each data source

Data Access and Discovery

- Open API's for both HydroServer and HydroCatalog
  - Possible to create independent data access client
- Most commonly used tools are:
  - HydroExcel
  - HydroDesktop

HydroExcel

- Uses Microsoft Excel to retrieve data from HydroServers



For Help Scroll Left

Specify the web service that will be used in all worksheets

WSDL Location: http://hydro1.sci.gsfc.nasa.gov/daac-bin/his/1.0/NLDAS_FORA_002.cgi?WSDL

Get Capabilities | Open Service Web Page | Get Sites | Get Variables | Get Series Catalog | Get All | Clear All

Below, results from Get Capabilities indicate which worksheets should work with the selected Web Service.

Sites	TRUE
Variables	TRUE
Site Info	TRUE
Series Catalog	TRUE
Time Series	TRUE

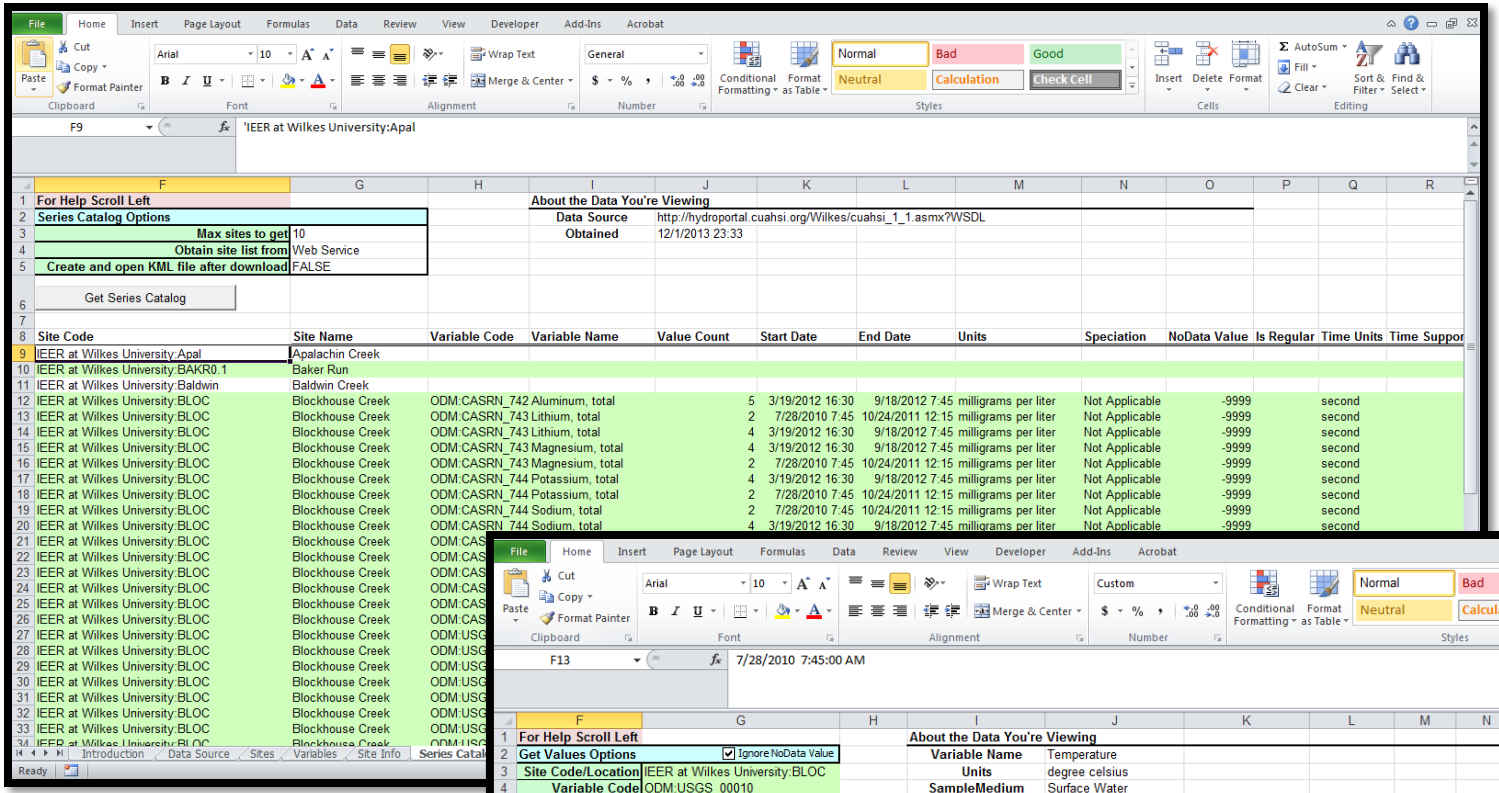
Example WaterOneFlow Web Services — Get Registered Services from HIS Central

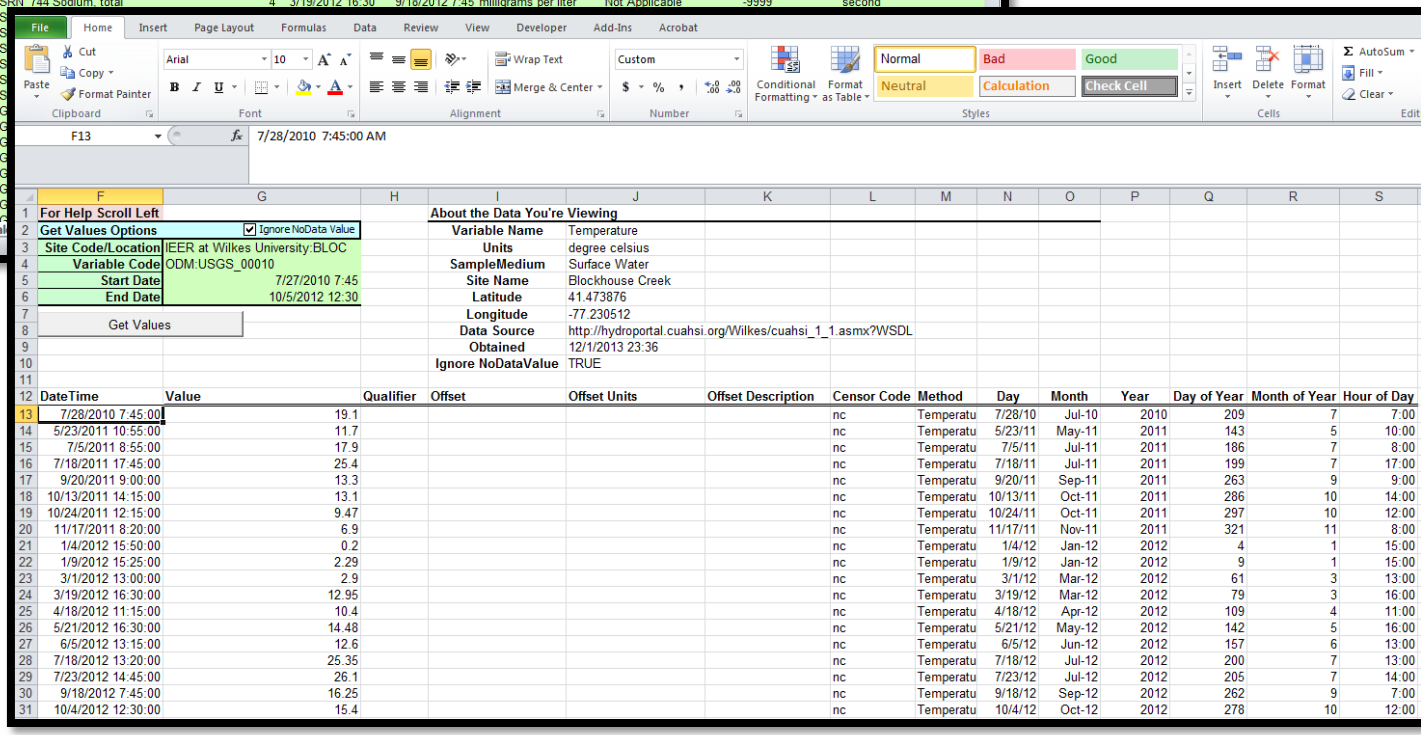
Organization	WSDL Location	Title	Site Count	Abstract	Value Count	West Longitude	South Latitude	East Long
Aalto university	http://peru.hydrodata.org/codehydro/services/cuahsi_1_1.asmx?wsdl	OGIMET - Peru	37	Global daily	2804961	-81.23278	-18.05	-69
Aalto university department of environmental ir	http://hydro.keskari.webfactional.com/services/cuahsi_1_1.php?wsdl	Lahti Urban Water Laboratory	3	The urban w	75	25.60947	60.99117	2
Aalto university department of environmental s	http://ogimet.hydrodata.org/codehydro/services/cuahsi_1_1.asmx?wsdl	fmi-test	64	OGIMET dat	91110	19.13278	59.5	30
Aalto university department of environmental s	http://ogimet.hydrodata.org/services/cuahsi_1_1.asmx?wsdl	OGIMET Meteorological Data - Europe	450	OGIMET glo	36698340	-57.89972	-66.66695	14
Aalto university department of environmental s	http://arctic.hydrodata.org/services/cuahsi_1_1.asmx?wsdl	OGIMET Arctic Climate Data	79	GIMET globa	4897100	-73	59.98278	-13
Baltimore Ecosystem Study	http://hydro2.umbc.edu/BESOD/cuahsi_1_0.asmx?WSDL	Baltimore Ecosystem Study Stream Chemistry Data	11		21486	-76.81722	39.27083	-76
Baltimore Ecosystem Study	http://hydro2.umbc.edu/BESSoil/cuahsi_1_0.asmx?WSDL	Baltimore Ecosystem Study Soils Data	28		140490	-76.82363	39.23514	-76
Boise State University, Civil Engineering	http://icewater.boisestate.edu/srbdataservices/cuahsi_1_0.asmx?WSDL	Snake River Basin, Modeled Streamflow	1	Preliminary	32872	-111.905	43.8258	-1
Boise State University, Hydrologic Sciences D	http://icewater.boisestate.edu/dcew2dataservices/cuahsi_1_0.asmx?WSDL	Dry Creek Experimental Watershed, SW Idaho	100	Dry Creek E	8763139	-116.1786	43.68834	-1
byu	http://hydrodata.netau.net/services/cuahsi_1_1.php?wsdl	hydrodata	0	Rainfall and	0	0	0	
Can Tho University	http://192.168.1.5/AnGiang_HydrologyData/cuahsi_1_1.asmx	AnGiang_HydrologyData	0	hydrology da	0	0	0	
Can Tho University, Vietnam, Wetlands Univer	http://hydro10.sdsc.edu/cantho/cuahsi_1_1.asmx?WSDL	Can Tho University Salinity Study	76	This study is	2119	104.8885	8.827816	10
Charles University Prague	http://hydrodata.info/webservices/cuahsi_1_1.asmx?WSDL	Central European Climate Data	354	The quality c	6387509	6.05	46	
Chesapeake Bay Environmental Observatory	http://eddy.ccny.cuny.edu/BENTHIC/cuahsi_1_1.asmx?WSDL	Benthic Data in Chesapeake Bay	4367		299004	-77.3197	36.4554	-7
Chesapeake Bay Information Management Sy	http://eddy.ccny.cuny.edu/CIMS/cuahsi_1_1.asmx?WSDL	Chesapeake Bay Information Management System	894	The Chesap	5418382	-79.89764	36.73611	
CIMA Research Foundation	http://hydro10.sdsc.edu/cima1/cuahsi_1_1.asmx?WSDL	CIMA1	839	Historical da	239115	6.68583	43.55798	10
Coweeta Hydrologic Laboratory and LTER	http://river.sdsc.edu/WaterOneFlow/Coweeta/cuahsi_1_1.asmx?WSDL	Coweeta Hydrologic Laboratory Public Data	2	A portion of	129663	-83.436	35.05116	-83
Crown of the Continent Observatory (Montana	http://his03.geol.umt.edu/COTCsnow/cuahsi_1_0.asmx?WSDL	Crown of the Continent Observatory Snow	2		56601	-113.753	48.62614	-1
CUAHSI	http://hydroportal.cuahsi.org/JPTEST2/cuahsi_1_1.asmx?WSDL	CUAHSI Test 2	1	This is a CU	10	-70.52422	41.38922	-70
CUAHSI	http://hydroportal.cuahsi.org/JPTEST3/cuahsi_1_1.asmx?WSDL	CUAHSI Test 3	1	This is a CU	10	-70.52422	41.38922	-70

Introduction | Data Source | Sites | Variables | Site Info | Series Catalog | Site Summary | Time Series | Statistics and Charts

Ready

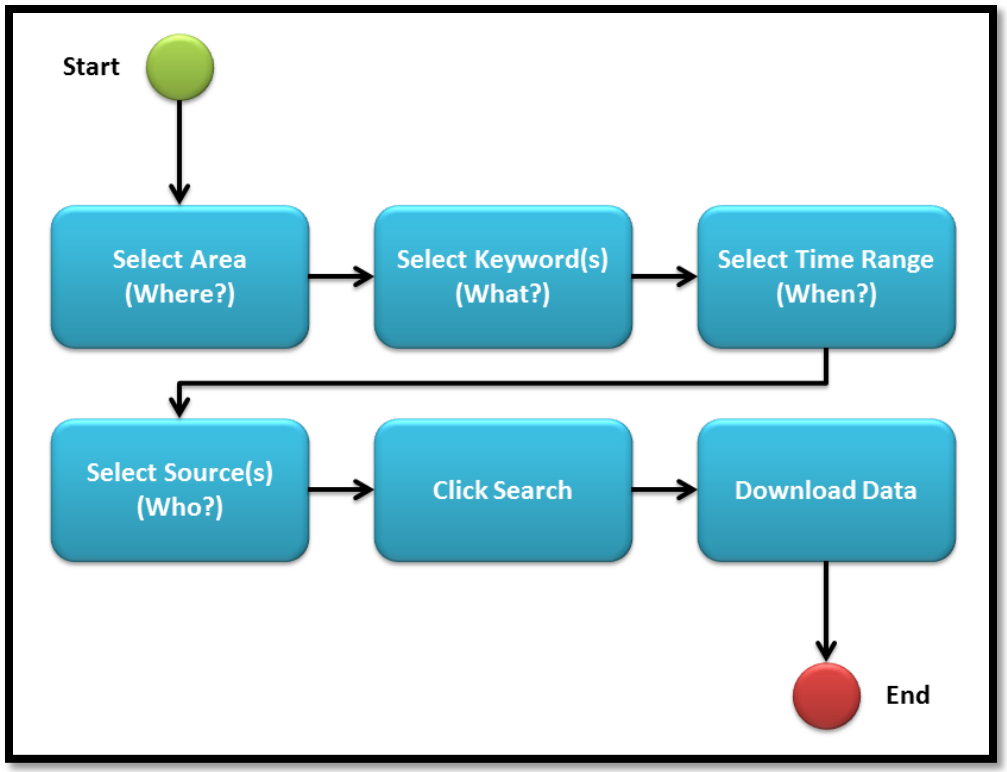
HydroExcel


F9
'IEER at Wilkes University:Apal
For Help Scroll Left
Series Catalog Options
Max sites to get 10
Obtain site list from Web Service
Create and open KML file after download FALSE
Get Series Catalog
About the Data You're Viewing
Data Source http://hydroportal.cuahsi.org/Wilkes/cuahsi_1_1.asmx?WSDL
Obtained 12/1/2013 23:33
Site Code | Site Name | Variable Code | Variable Name | Value Count | Start Date | End Date | Units | Speciation | NoData Value | Is Regular | Time Units
IEER at Wilkes University:Apal | Apalachin Creek
IEER at Wilkes University:BAKR0.1 | Baker Run
IEER at Wilkes University:Baldwin | Baldwin Creek
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_742 | Aluminum, total | 5 | 3/19/2012 16:30 | 9/18/2012 7:45 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_743 | Lithium, total | 2 | 7/28/2010 7:45 | 10/24/2011 12:15 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_743 | Lithium, total | 4 | 3/19/2012 16:30 | 9/18/2012 7:45 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_743 | Magnesium, total | 4 | 3/19/2012 16:30 | 9/18/2012 7:45 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_743 | Magnesium, total | 2 | 7/28/2010 7:45 | 10/24/2011 12:15 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_744 | Potassium, total | 4 | 3/19/2012 16:30 | 9/18/2012 7:45 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_744 | Potassium, total | 2 | 7/28/2010 7:45 | 10/24/2011 12:15 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_744 | Sodium, total | 2 | 7/28/2010 7:45 | 10/24/2011 12:15 | milligrams per liter | Not Applicable | -9999 | second
IEER at Wilkes University:BLOC | Blockhouse Creek | ODM:CASRN_744 | Sodium, total | 4 | 3/19/2012 16:30 | 9/18/2012 7:45 | milligrams per liter | Not Applicable | -9999 | second
Introduction | Data Source | Sites | Variables | Site Info | Series Catal
Ready

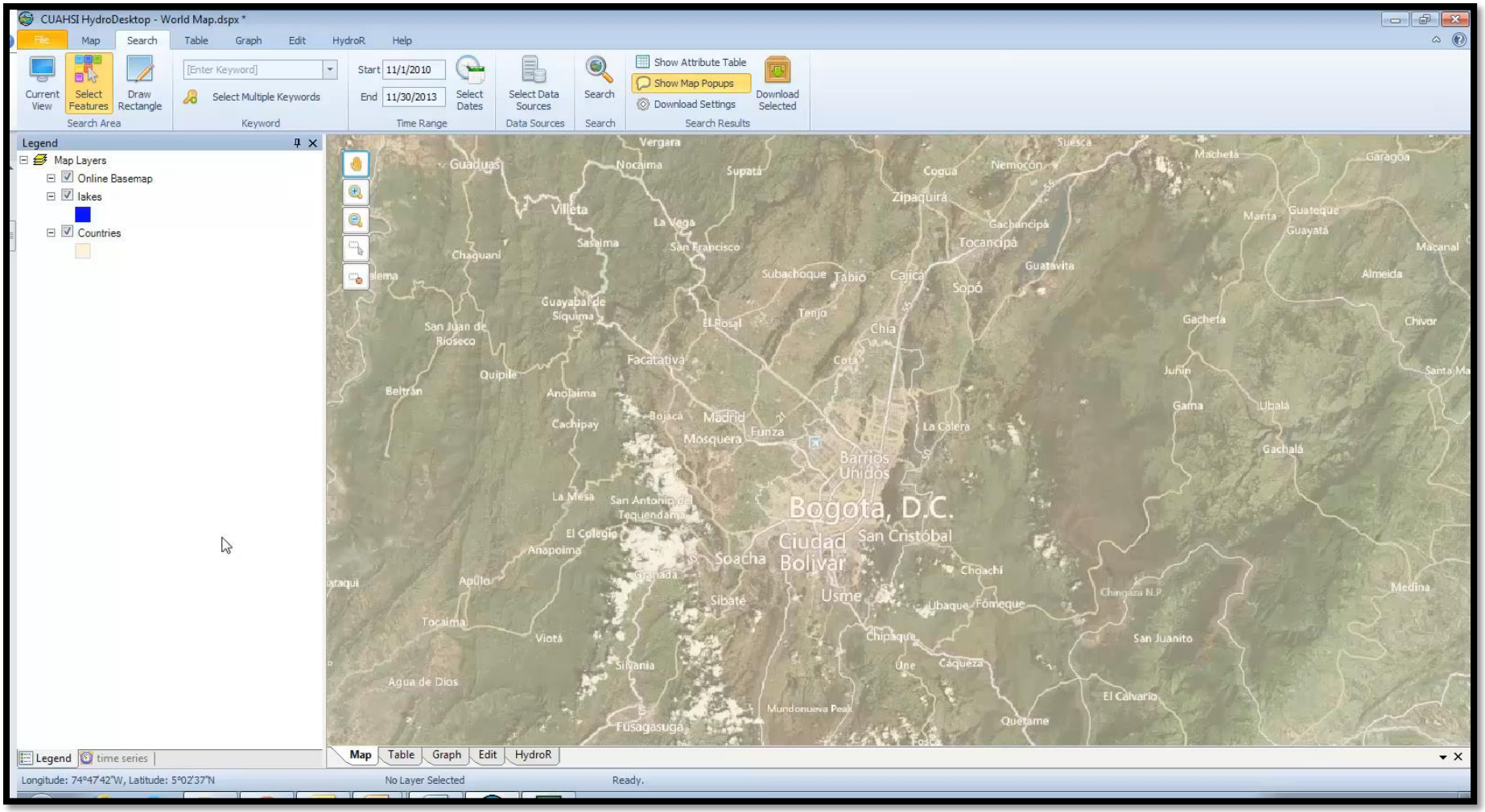

F13
7/28/2010 7:45:00 AM
For Help Scroll Left
Get Values Options
Ignore NoData Value
Site Code/Location IEER at Wilkes University:BLOC
Variable Code ODM:USGS_00010
Start Date 7/27/2010 7:45
End Date 10/5/2012 12:30
Get Values
About the Data You're Viewing
Variable Name Temperature
Units degree celsius
SampleMedium Surface Water
Site Name Blockhouse Creek
Latitude 41.473876
Longitude -77.230512
Data Source http://hydroportal.cuahsi.org/Wilkes/cuahsi_1_1.asmx?WSDL
Obtained 12/1/2013 23:36
Ignore NoDataValue TRUE
DateTime | Value | Qualifier | Offset | Offset Units | Offset Description | Censor Code | Method | Day | Month | Year | Day of Year | Month of Year | Hour of Day
7/28/2010 7:45:00 | 19.1 | nc | Temperatu | 7/28/10 | Jul-10 | 2010 | 209 | 7 | 7:00
5/23/2011 10:55:00 | 11.7 | nc | Temperatu | 5/23/11 | May-11 | 2011 | 143 | 5 | 10:00
7/5/2011 8:55:00 | 17.9 | nc | Temperatu | 7/5/11 | Jul-11 | 2011 | 186 | 7 | 8:00
7/18/2011 17:45:00 | 25.4 | nc | Temperatu | 7/18/11 | Jul-11 | 2011 | 199 | 7 | 17:00
9/20/2011 9:00:00 | 13.3 | nc | Temperatu | 9/20/11 | Sep-11 | 2011 | 263 | 9 | 9:00
10/13/2011 14:15:00 | 13.1 | nc | Temperatu | 10/13/11 | Oct-11 | 2011 | 286 | 10 | 14:00
10/24/2011 12:15:00 | 9.47 | nc | Temperatu | 10/24/11 | Oct-11 | 2011 | 297 | 10 | 12:00
11/17/2011 8:20:00 | 6.9 | nc | Temperatu | 11/17/11 | Nov-11 | 2011 | 321 | 11 | 8:00
1/4/2012 15:50:00 | 0.2 | nc | Temperatu | 1/4/12 | Jan-12 | 2012 | 4 | 1 | 15:00
1/9/2012 15:25:00 | 2.29 | nc | Temperatu | 1/9/12 | Jan-12 | 2012 | 9 | 1 | 15:00
3/1/2012 13:00:00 | 2.9 | nc | Temperatu | 3/1/12 | Mar-12 | 2012 | 61 | 3 | 13:00
3/19/2012 16:30:00 | 12.95 | nc | Temperatu | 3/19/12 | Mar-12 | 2012 | 79 | 3 | 16:00
4/18/2012 11:15:00 | 10.4 | nc | Temperatu | 4/18/12 | Apr-12 | 2012 | 109 | 4 | 11:00
5/21/2012 16:30:00 | 14.48 | nc | Temperatu | 5/21/12 | May-12 | 2012 | 142 | 5 | 16:00
6/5/2012 13:15:00 | 12.6 | nc | Temperatu | 6/5/12 | Jun-12 | 2012 | 157 | 6 | 13:00
7/18/2012 13:20:00 | 25.35 | nc | Temperatu | 7/18/12 | Jul-12 | 2012 | 200 | 7 | 13:00
7/23/2012 14:45:00 | 26.1 | nc | Temperatu | 7/23/12 | Jul-12 | 2012 | 205 | 7 | 14:00
9/18/2012 7:45:00 | 16.25 | nc | Temperatu | 9/18/12 | Sep-12 | 2012 | 262 | 9 | 7:00
10/4/2012 12:30:00 | 15.4 | nc | Temperatu | 10/4/12 | Oct-12 | 2012 | 278 | 10 | 12:00

HYDRODESKTOP

- Open source Geographic Information System (GIS)
  - Extensible → Developers can create “plugins” that enhance functionality
- Searches HydroCatalog for data; downloads data from HydroServers



HYDRODESKTOP



Year 1 of Becoming the CUAHSI Water Data Center

From “prototype” to “product”.

Case Studies of Data Publication

- Federal Agencies: USGS National Water Information System & NASA data products
- Shale Network
- Ipswich River Watershed Association
- Little Bear River Experimental Watershed

United States Geological Survey (USGS)

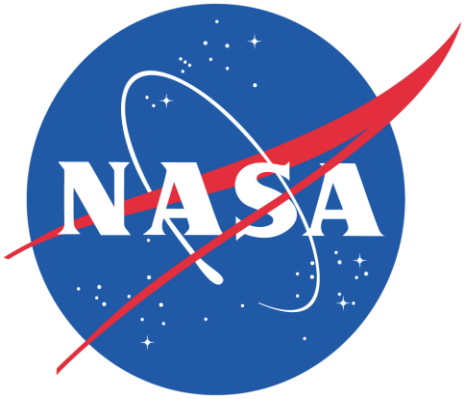
- U.S. Federal Agency that monitors ecosystems and environmental health
- National Water Information System is a monitoring system with about 1.5 million sites across the United States

NWIS Data in the CUAHSI HIS Catalog

- Data are transformed by CUAHSI HIS team
  - Scripts written in XSLT transform metadata to comply with Observations Data Model
- Largest volume of data in CUAHSI Water Data Catalog
  - Trusted source of water data
  - Access to this source important for community adoption of data system

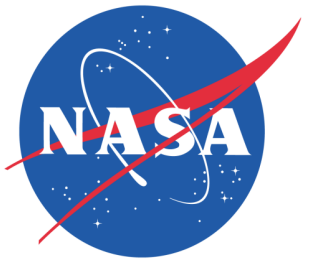
National Aeronautics and Space Administration (NASA)

- NASA operates US civilian Space programs
- NASA Goddard Earth Sciences (GES) Data and Information Services Center publishes a number of data products created from models and information obtained by satellites



NASA Data Products

- Gridded data, provided in a format more easily accessible to hydrologists
- High community demand
  - Global coverage
  - Soil moisture, precipitation, temperature, wind speed, and more
- Data are transformed by government with support from CUAHSI HIS team



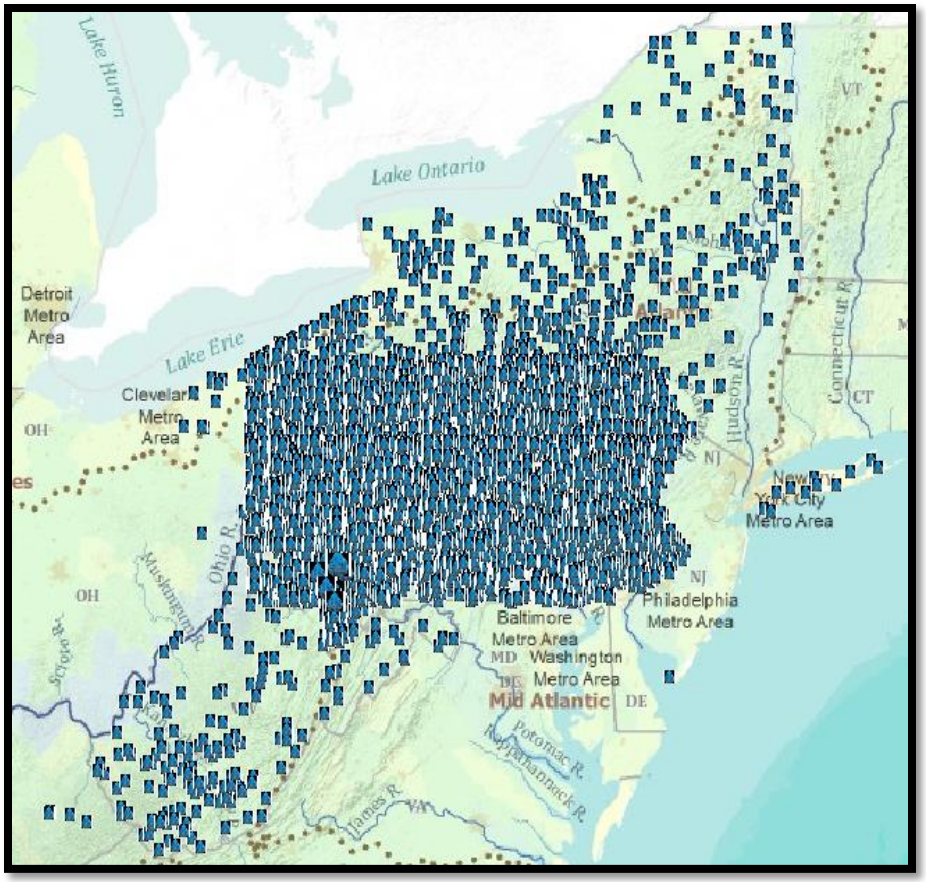
Shale Network

- Research project funded by National Science Foundation
- Investigating possible impacts of exploitation of Shale gas on water resources
- Led by Professor Susan Brantley, Pennsylvania State University with…
  - Pittsburgh University
  - Dickinson College
  - CUAHSI

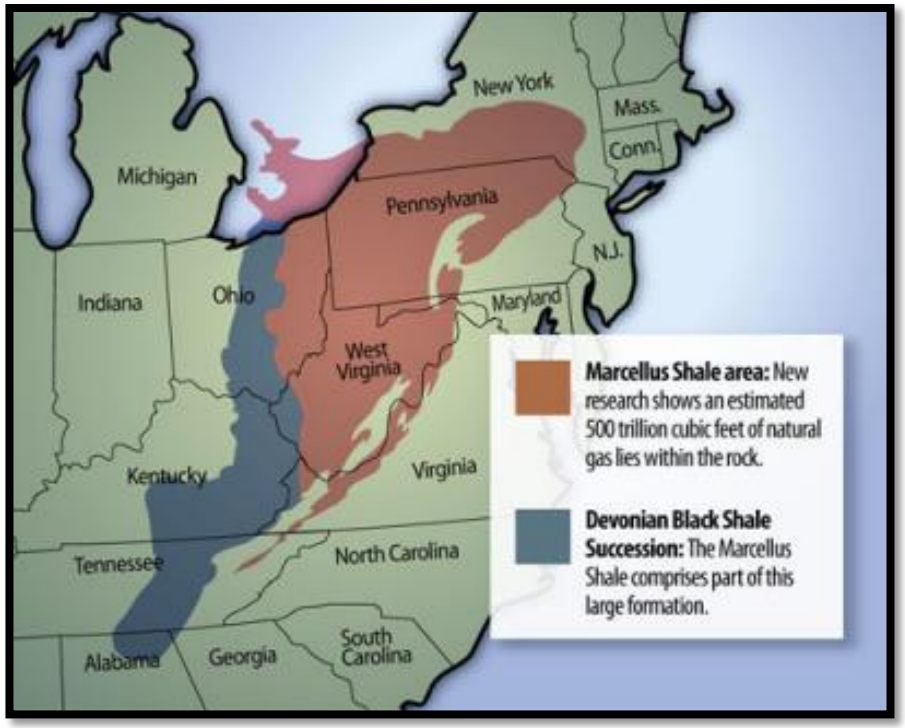
Shale Network

- Aggregating data from multiple sources including
  - Academic research projects
  - Local governments
  - Citizen scientists
  - Want to include data from industry
- CUAHSI supports project through user support and data hosting

Shale Network



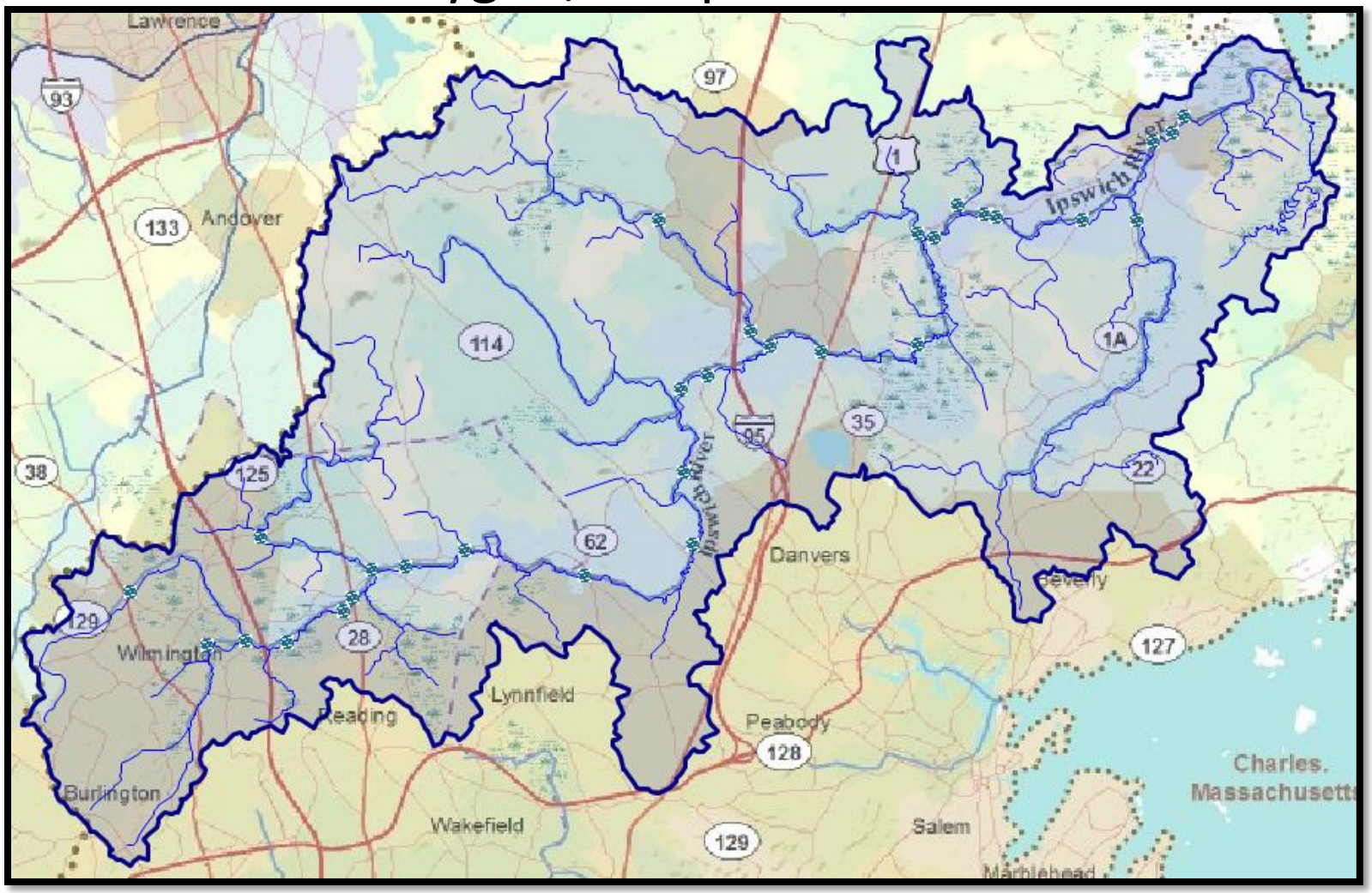
Locations of observations in the Shale Network database as seen in HydroDesktop



Above Image Source: Buffalo Business First

Ipswich River Watershed Association

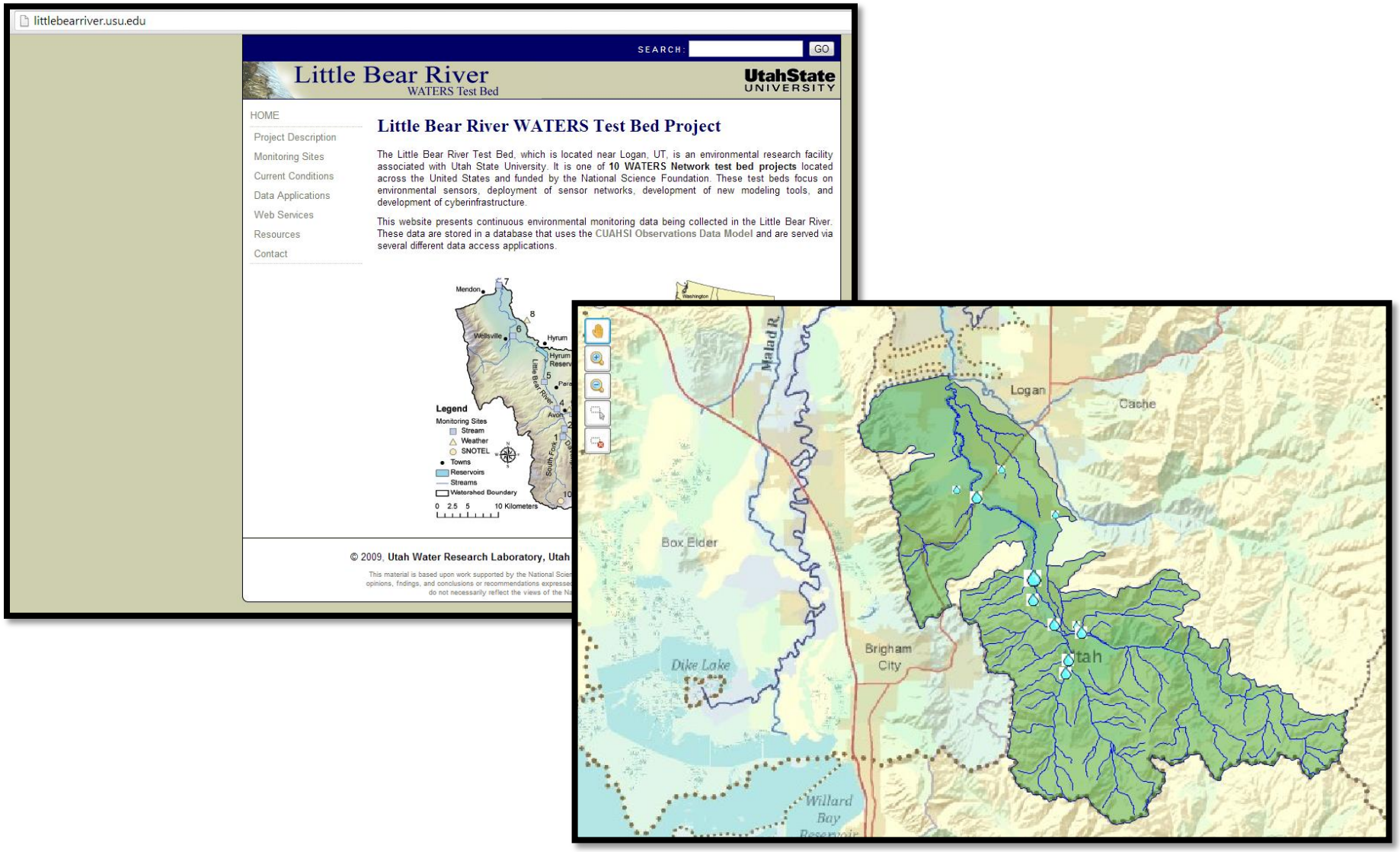
- Citizen science group that monitors local river
  - Dissolved oxygen, temperature



Little Bear River Experimental Watershed

- Monitoring of local watershed by researchers at Utah State University
- Studying short-term variability in discharge, water quality, and pollutant loading within the Little Bear River, which drains a mixed agricultural, range, and forested watershed in northern Utah, USA

Little Bear River Experimental Watershed



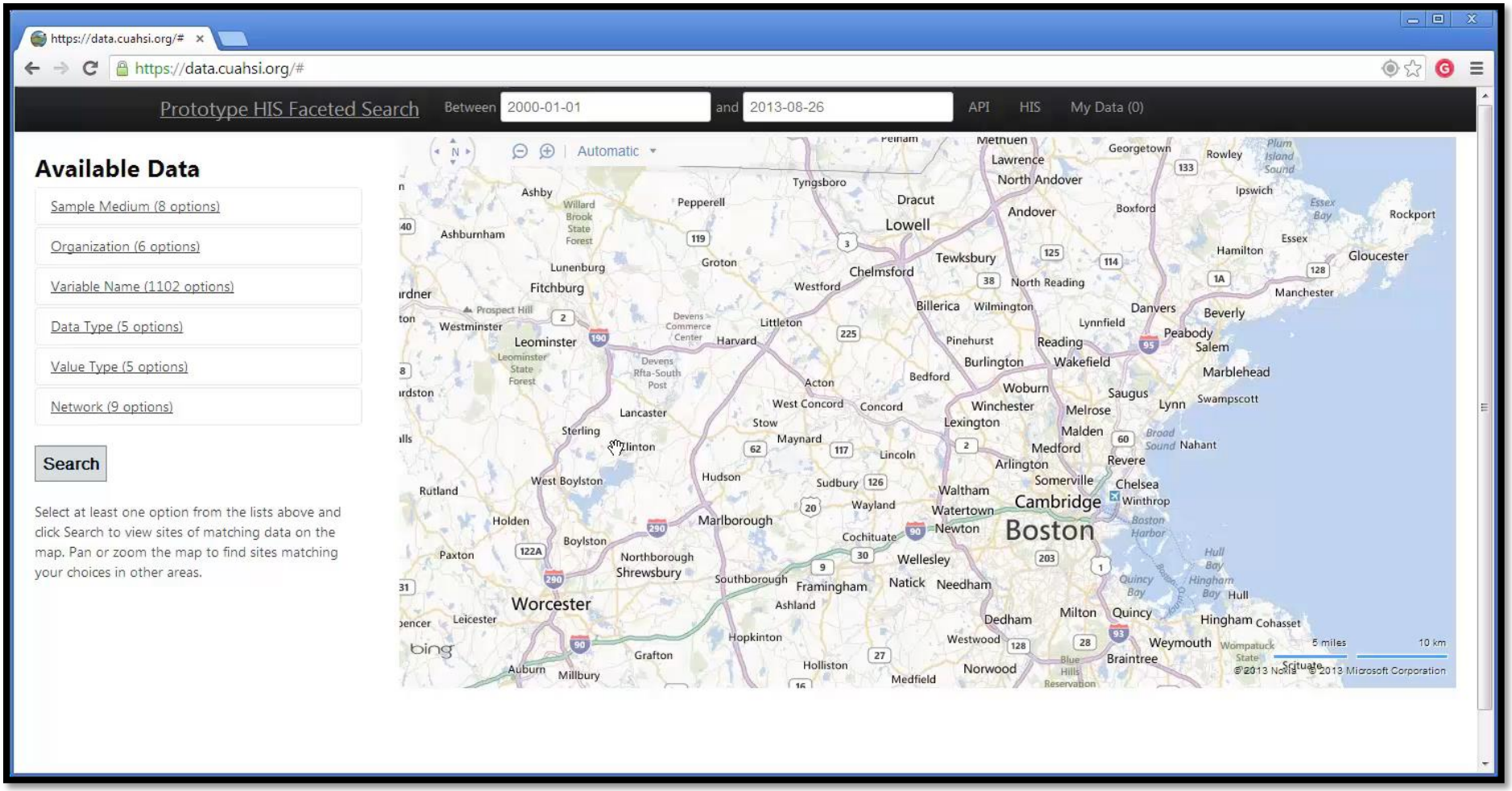
Near Future Goals

1. New data discovery client
2. Easing data publication process
3. Federated Catalogs

Faceted Search

- New method for data discovery
- Web-based application that will utilize cloud capabilities for faster search returns
- Web-based application that filters search results based upon metadata fields

Faceted Search Client



Challenges in Data Publication Process

- Need Improved Semantics
  - Integrate other disciplines via linked data (SKOS)
- Need to Re-Engineer Data Uploader
  - Improved usability → Develop feedback mechanism for data that does not comply with ODM specifications

Federated Catalogs

- Create a “catalog of catalogs”
- International interest in deploying CUAHSI HIS software
- Global Earth Observation System of Systems: Infrastructure on the global scale for comprehensive, near-real-time environmental data

Long-term Questions for the CUAHSI WDC

- How to create persistent identifiers for data sources that change in real time?
- What kinds of non-time-series data to make available within the data center?
- What kinds of data fusion services best enable research?

Questions?

CUAHSI

UNIVERSITIES ALLIED FOR WATER RESEARCH

Jon Pollak
jpollak@cuahsi.org