

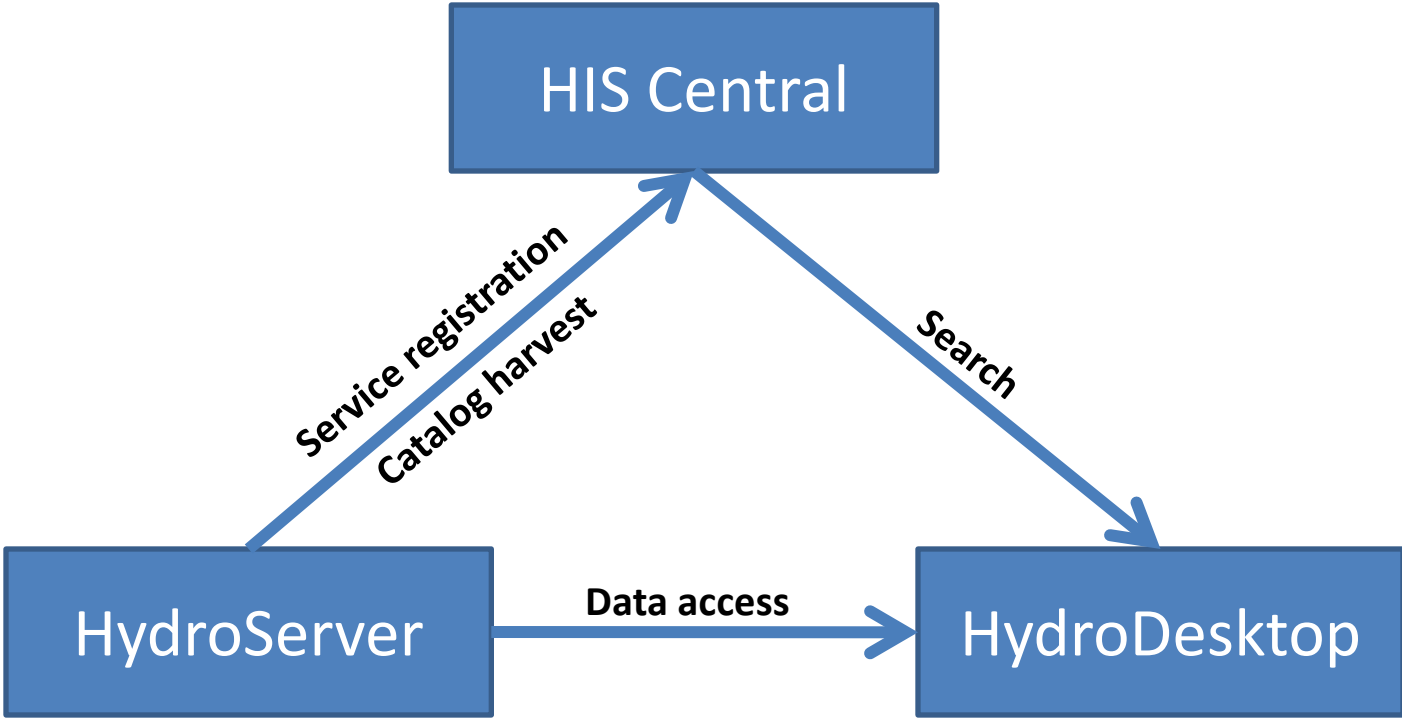
Implementing a Community-governed Ontology: Experiences from the Water Sciences Community

Rick Hooper (CUAHSI)

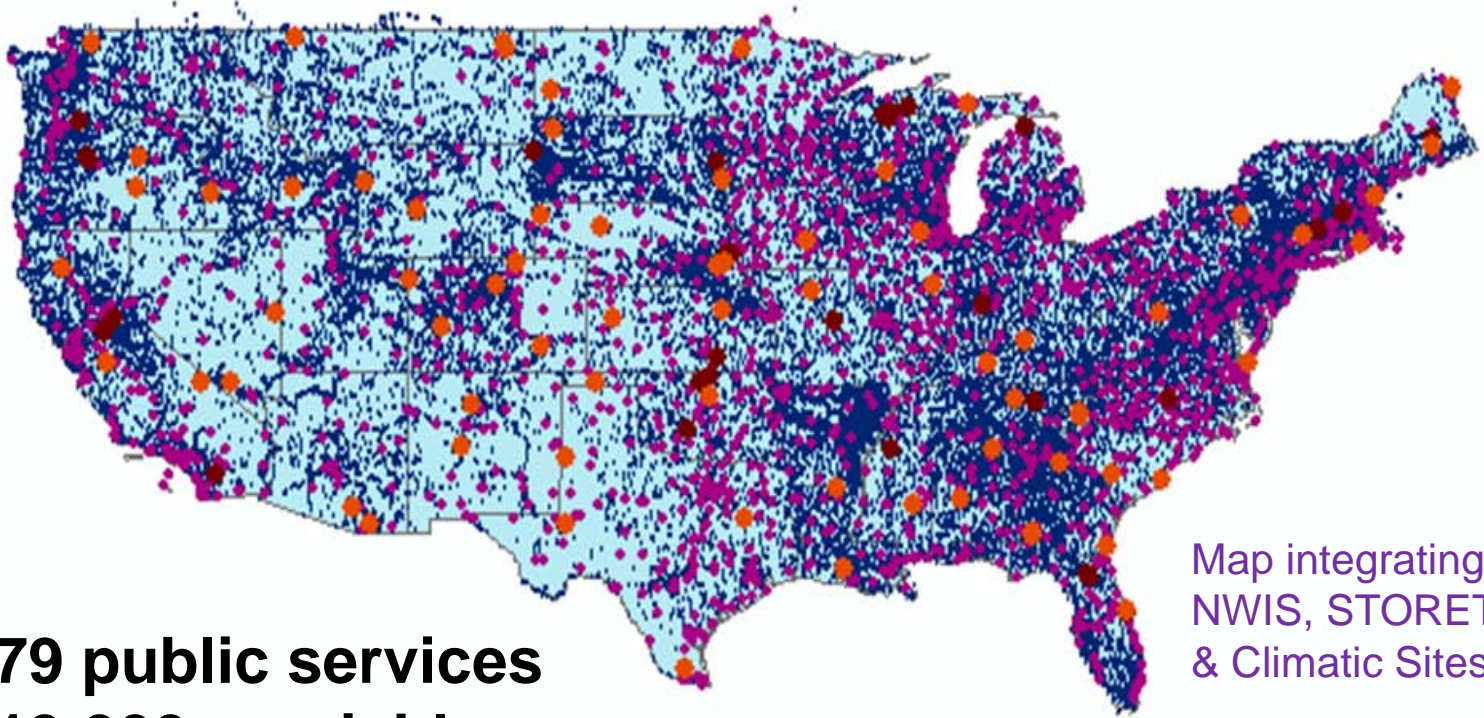
Michael Piasecki (CCNY)

Ilya Zaslavksy and David Valentine (SDSC)

Services-Oriented Architecture for Water Data



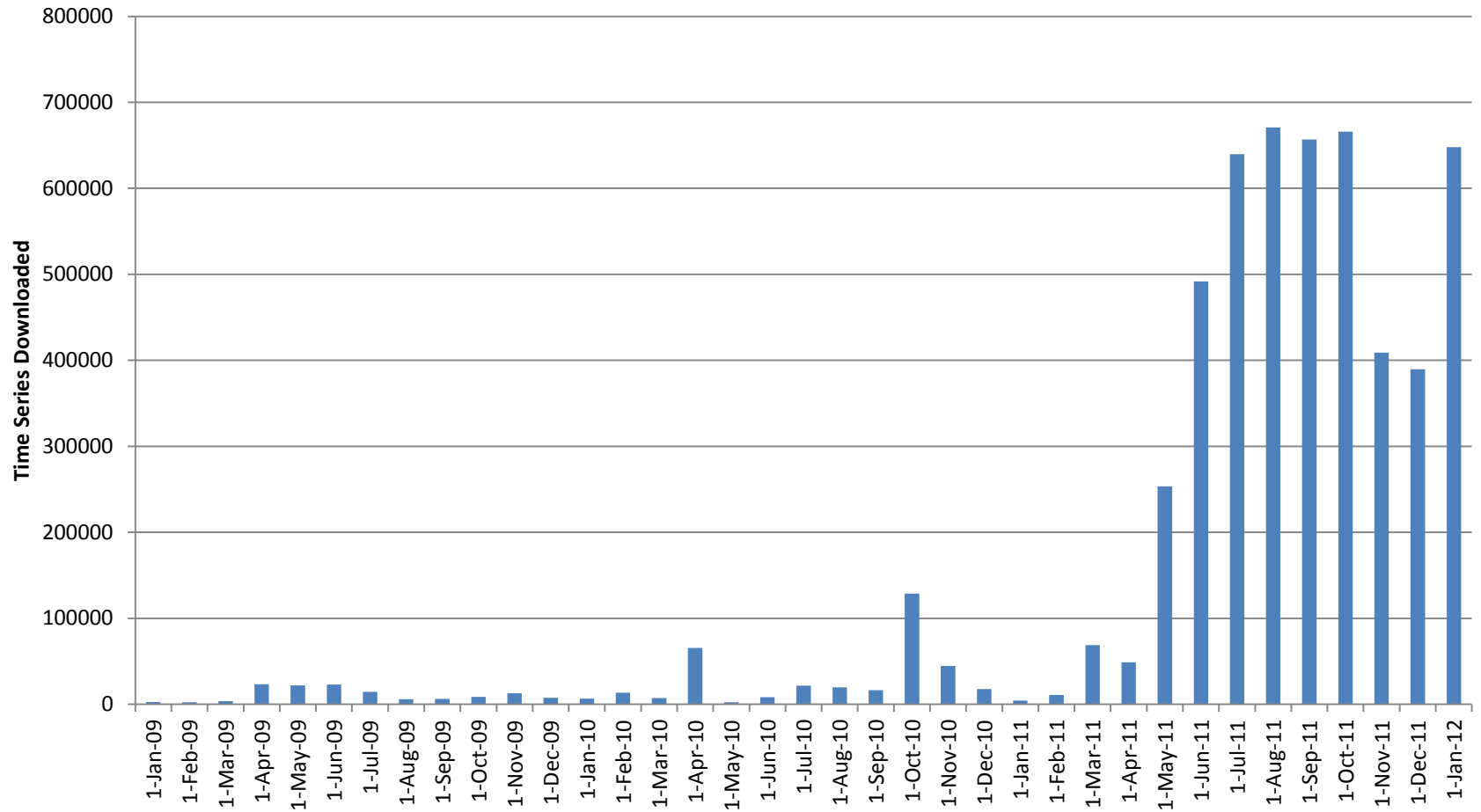
Metadata Catalog, Feb 2012



Map integrating
NWIS, STORET,
& Climatic Sites

79 public services
13,000+ variables
2.3+ million sites
23.3 million series
Referencing 100+ billion data values

HIS Usage



For more on the HIS Story

his.cuahsi.org

The screenshot displays the CUAHSI HIS website interface. At the top left is the logo with the text "CUAHSI HIS Sharing hydrologic data". To the right is a descriptive paragraph: "CUAHSI's Hydrologic Information System (CUAHSI-HIS) provides web services, tools, standards and procedures that enhance access to more and better data for hydrologic analysis." Below the logo is a navigation menu with links: Home, How To, Components, Community, Publications, About Us, and Contact Us. A search bar is located on the right side of the menu.

The main content area features a paragraph: "The CUAHSI Hydrologic Information System (HIS) is an internet-based system for sharing hydrologic data. It is comprised of databases and servers, connected through web services, to client applications, allowing for the publication, discovery and access of data."

Below this is a diagram titled "Key Components of CUAHSI-HIS:"

```
graph TD; HS[HydroServer Data Publication] -- "Metadata Services  
Service Registration  
and Catalog Harvesting" --> HC[HIS Central Data Discovery]; HC -- "Search Services  
Geographic, Semantic, Time  
and Network Search" --> HD[HydroDesktop and other clients Data Access]; HS -- "Data Services  
Water and Spatial Data" --> HD;
```

On the right side of the page, there are sections for "Quick Links" and "What's New".

Quick Links

- HydroDesktop
- HydroExcel
- HydroGet
- FetchWaterML
- WaterML Web Services
- ODM Database
- HydroServer
- Master Controlled Vocabulary
- HydroTagger
- HIS Central

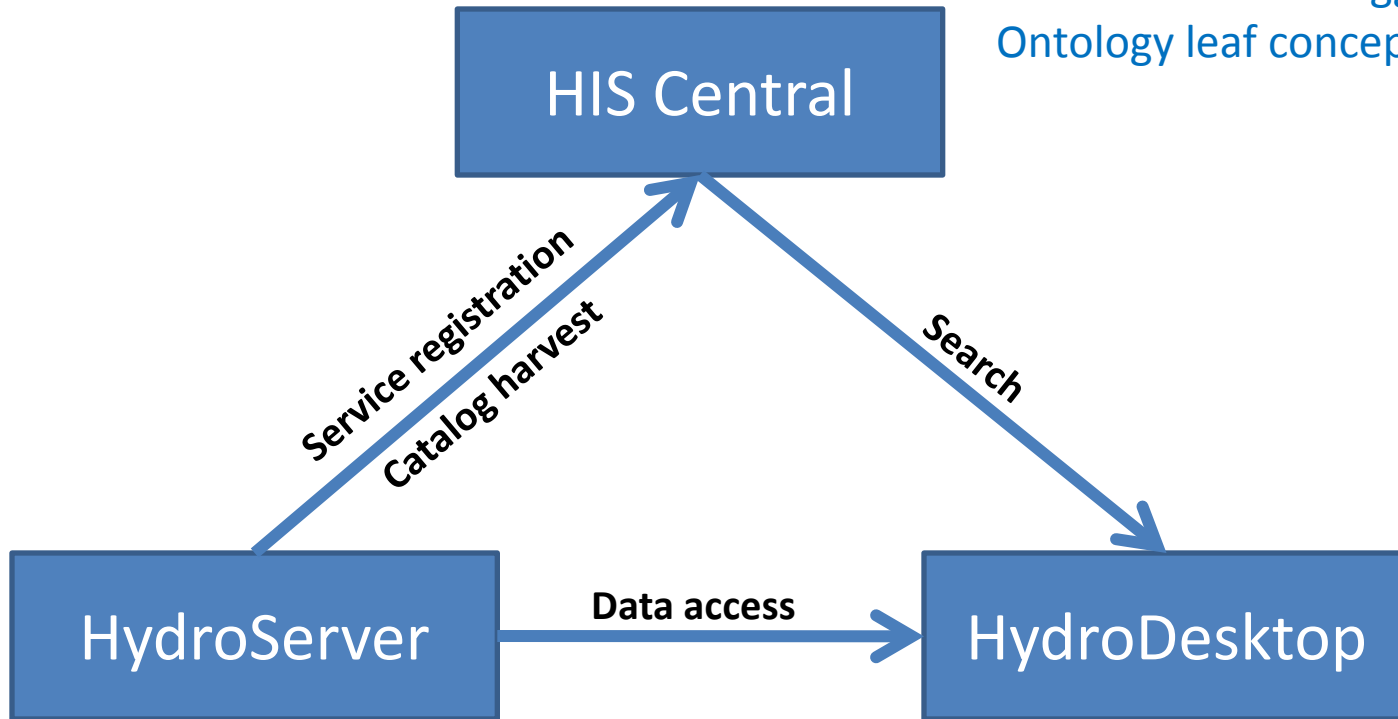
What's New

- Hydrologic Data Needs Survey**
Help us help you, by filling out our Hydrologic Data Publication, Discovery, Access and Analysis Needs Survey. This survey follows on to those conducted in 2004 and 2006. Results of these surveys have had a significant influence on the direction of this project.
- Hope to see you at the Water Data Service Workshop, July 21-22, 2010**
This workshop is an introduction to the CUAHSI HIS web service approach to sharing and accessing hydrologic data, including the key tools and technologies: HydroServer (store and publish), HIS Central (catalog and search), and HydroDesktop (access and analyze). The workshop follows the CUAHSI Biennial Colloquium in Boulder, CO.
- HIS at the AWRA Spring Specialty Conference: GIS & Water Resources VI**
The keynote by ESRI's Jack Dangermond described how ESRI is integrating HIS technology into their new products. There was also a number of other HIS related presentations, and we had demos and more at the CUAHSI booth at the conference on, March 29-31, 2008, and the workshops that followed it on April 1-2, in Orlando, FL.

At the bottom left of the screenshot, the word "Done" is visible.

Services-Oriented Architecture for Water Data

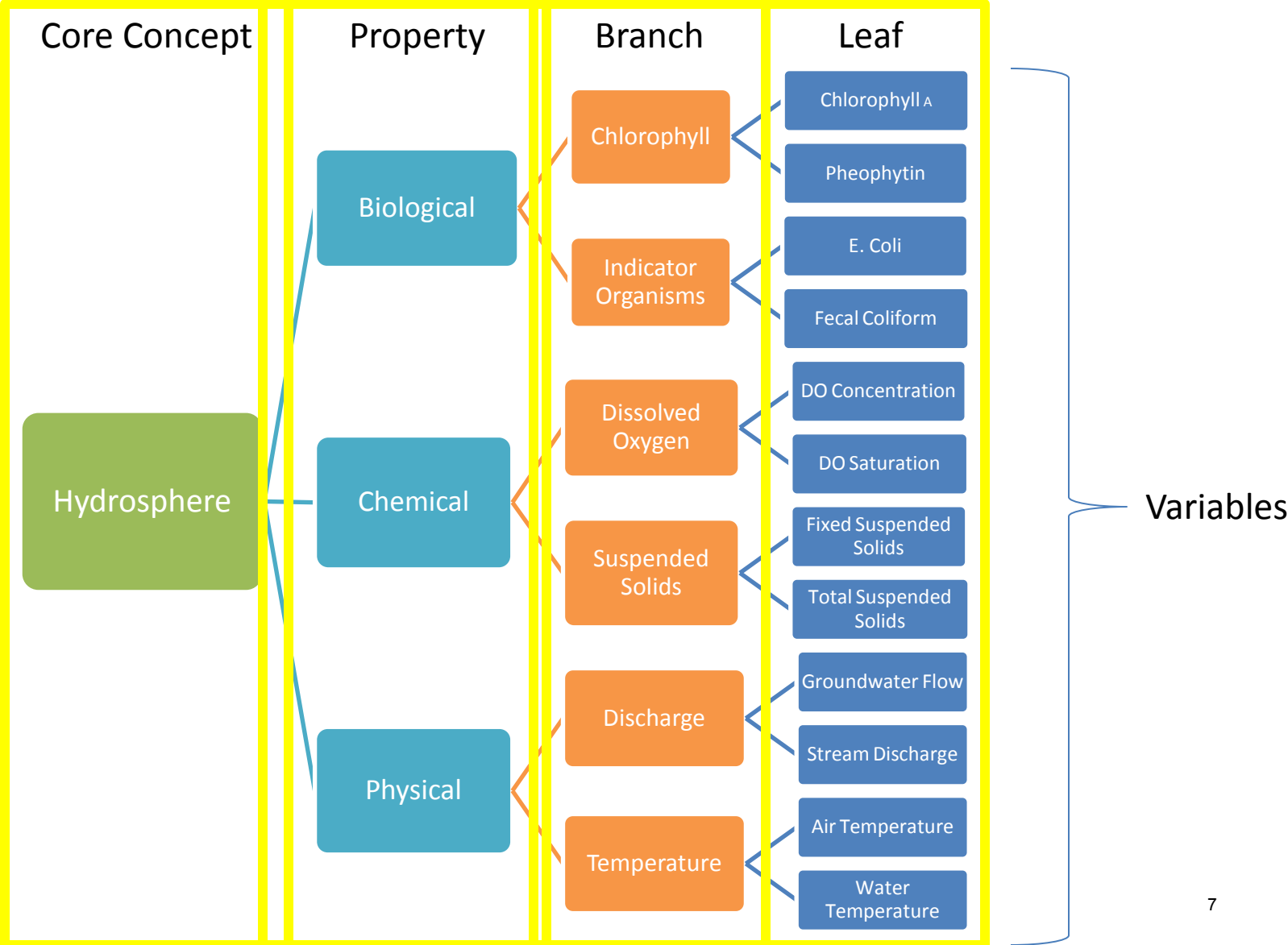
Variable Names “tagged” to
Ontology leaf concepts for search



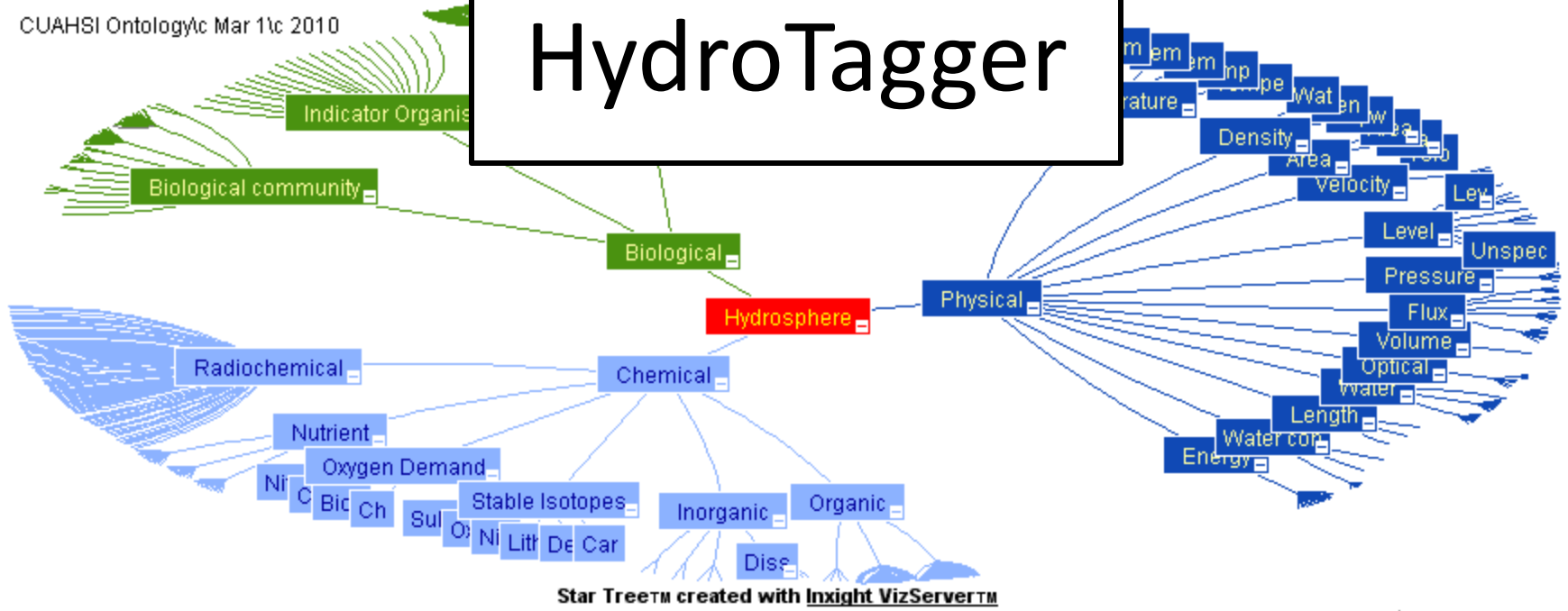
Controlled Vocabulary for Variable Name

Search for Variable Name,
Date/Time Box

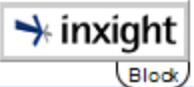
Ontology Taxonomy



HydroTagger



Star Tree™ created with Inxight VizServer™



Variable Name	Code	Medium	
sampling depth, feet	nwisuv:00003	unknown	select
sample accounting number	nwisuv:00008	unknown	select
location in cross sectio...	nwisuv:00009	unknown	select
temperature, water, degr...	nwisuv:00010	unknown	select
temperature, water, degr...	nwisuv:00011	unknown	select

1 2 3 4 5 6 7 8 9 10 ...

Variable:
Temperature, Water, Degre

Mapping:
Temperature, water

[Map!](#)

Variable	Keyword	
gage height, feet	water depth, stream	delete
discharge, cubic feet per second	discharge, stream	delete
barometric pressure, not corrected to sea level, millibars	atmospheric pressure	delete
acoustic signal strength, units		delete

Each **Variable** in your data is connected to a corresponding **Concept**

Is Taxonomy Important?

- No unique or 'correct' taxonomy
- Users more interested in controlled vocabulary and leaf concepts
- Any logical taxonomy is acceptable for navigation from broad to narrow term

Enabling Effective Searches

- Synonyms are important
 - “stream discharge” = “stream flow”
- Precision of return
 - Publishers want precise description
 - Need sufficient metadata on initial return to narrow search
 - Can one client serve expert and non-expert users?
 - ***Should leaf concepts be the variable names or broader than variable name?***
- Dimensionality of search
 - Variable/time/space isn't sufficient
 - Sample Medium, Site Type are likely first-order terms

Data and Metadata Models

- Sample (*ex situ*) metadata more complex than sensor (*in situ*) data
- Aquatic chemistry vs. solid phase chemistry example
- Legacy data systems overload different metadata fields to capture metadata
- Example: USGS NWIS includes medium and sample fraction in variable name
- Result: Need to reconcile data models among various data publishers
- ***Judgment call: How much metadata should system record and transmit? When should scientist consult original source?***

Ways Forward

- We can never record ‘sufficient’ metadata
- Enough metadata to avoid misinterpretation
- Easy access to ‘original’ source for more metadata
- Create ‘communities’ of data users to allow discussion of appropriate use
 - By Source (USGS), By Service (USGS IID database),
By Variable Name (Cadmium), By Time Series
(Cadmium concentration at Site 0110235)