
Microsoft® Research
Faculty Summit
10
YEAR ANNIVERSARY

Microsoft Research

Water for a Thirsty World: How Can Information Technology Help?

Part Two

Ilya Zaslavsky
San Diego Supercomputer Center
University of California, San Diego

Water Scarcity: A Deepening Problem

Microsoft Research

Drying and warming climate

Growing urban demand

Over-allocation to irrigation

The big

8

water scarcity factors

The environmental flows imperative

Uncapped groundwater extraction

Bushfire recovery impacts

Expanding plantations

Expanding farm dams

Slide Courtesy BOM, Australia

The Water Information Value Ladder

Microsoft Research

Data >>> Information >>> Insight

>>> Increasing value >>>

Forecasting

Reporting

Analysis

Integration

Distribution

Aggregation

Quality assurance

Collation

Monitoring

Done poorly

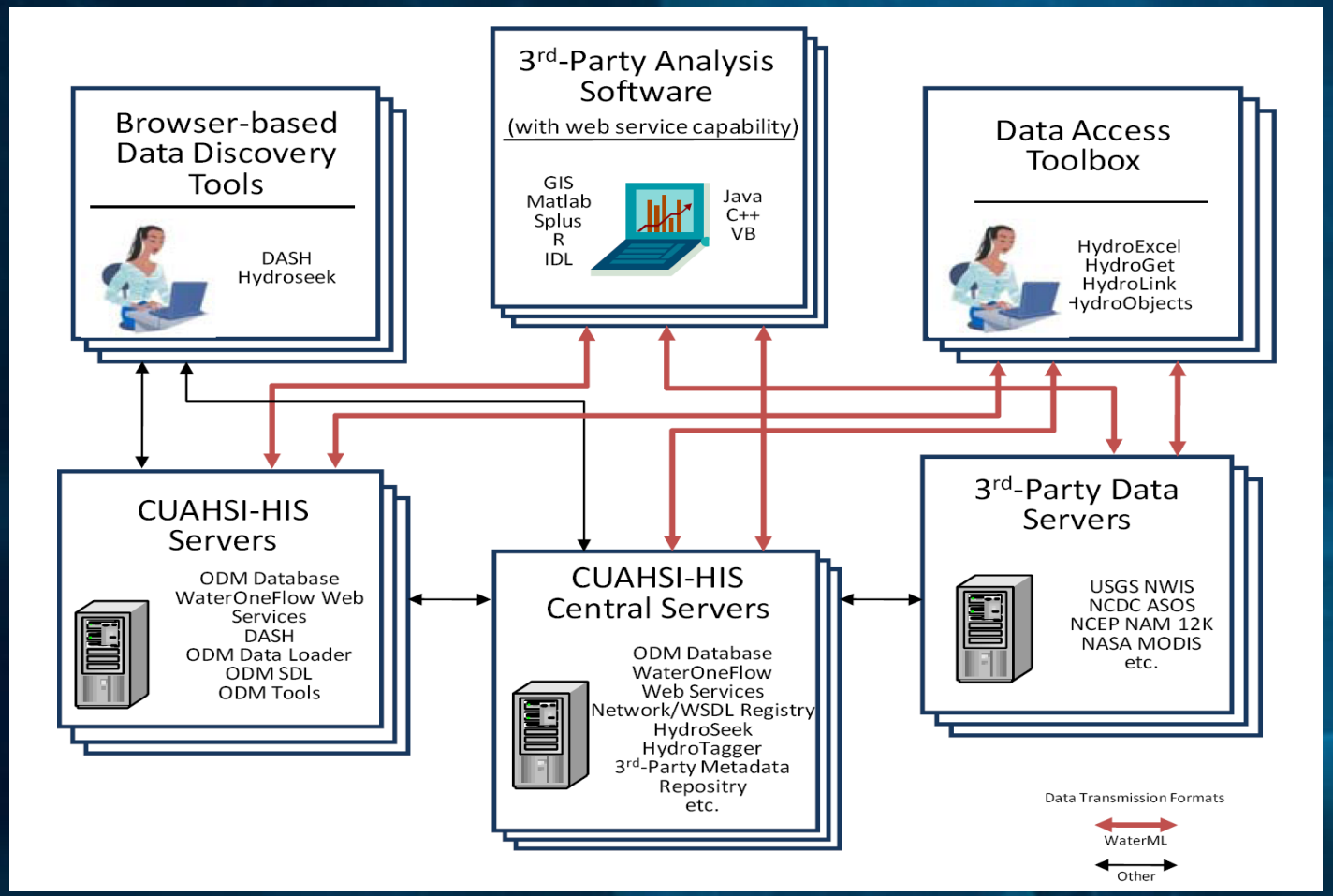
Done poorly to well

Generally done well, by many groups, but could be vastly improved with new IT

Slide Courtesy BOM, Australia

What is CUAHSI HIS?

Microsoft Research



CUAHSI (Consortium of Univ for the Advancement of Hydrologic Science, Inc.) Hydrologic Information System: NSF support through 2012 (GEO)

Partners:

Academic: 11 NSF hydrologic observatories, CEO:P projects, LTER…

Government: USGS, EPA, NCDC, NWS, state and local

Commercial: Microsoft, ESRI, Kisters

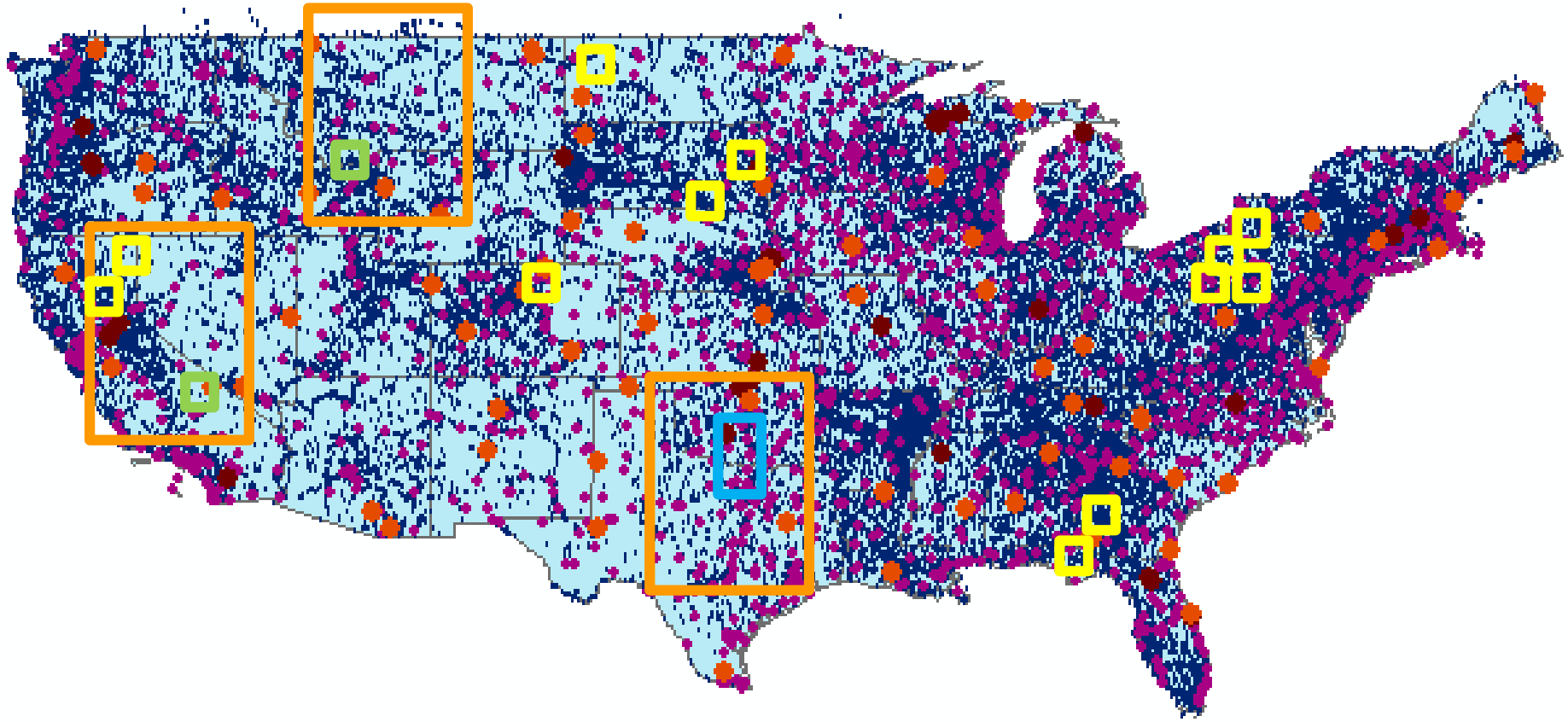
International: Australia, UK

Standardization: OGC, WMO (Hydrology Domain WG, CHy); adopted by USGS, NCDC

An online distributed system to support the sharing of hydrologic data from multiple repositories and databases via standard water data service protocols; software for data publication, discovery, access and integration.

Observation Stations Map for the US

Microsoft Research



Build a common window on water data using web services

How You Create an Information System Where…

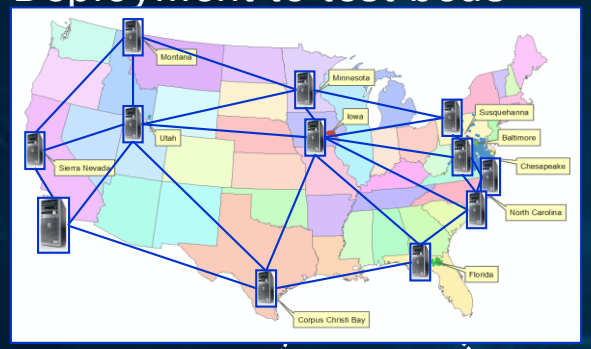
Microsoft Research

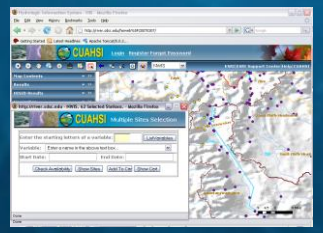
- There are N (widely distributed) sites that collect local data, possibly over time, organized in multiple observation networks
- Each site may collect a different set of data variables Vij
- Any Vij might come from sensor streams, local stores, compiled archives, differently aggregated and accessed…
- The data collection points N may be spatially related (e.g., in a river network, along a mountain ridge)
- Variables and values may be related through [evolving] ontologies
- Queries will have different spatiotemporal aggregate functions over data variables, and the spatial relationships across their sites
- The integration of data may need variable reconciliation and value reconciliation, spatio-temporal interpolation
- There are multiple user roles, research scenarios, models
- Other types of information (e.g. remote sensing, climate change projections) are generated at different spatial and temporal resolution, and need to be interoperable. Also: publications; multimedia; catalogs, web pages; policies and rules…
- … and there are millions of catalog records, and TBs of data

HIS Service Oriented Architecture

Microsoft Research

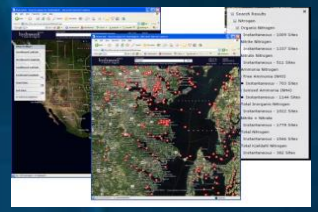
Deployment to test beds





Customizable web interface (DASH)

Global search



Other popular online clients

HTML - XML

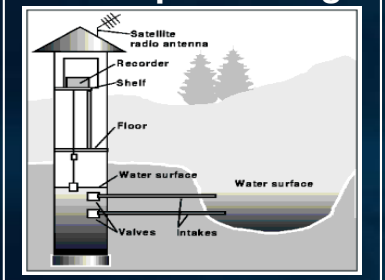
HIS Central Registry & Harvester

Water Data Web Services, WaterML

WSDL - SOAP

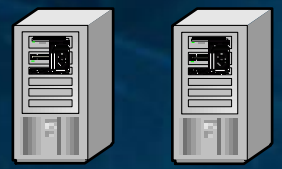
Metadata catalogs | **Ontology** | **Controlled vocabularies** | **ETL services**

Data publishing



- WSDL and ODM registration
- Ontology tagging (Hydrotagger)
- ODM DataLoader
- Streaming Data Loading
- QA/QC
- Server config tools

Test bed HIS Servers



HIS Lite Servers

Central HIS servers

External data providers

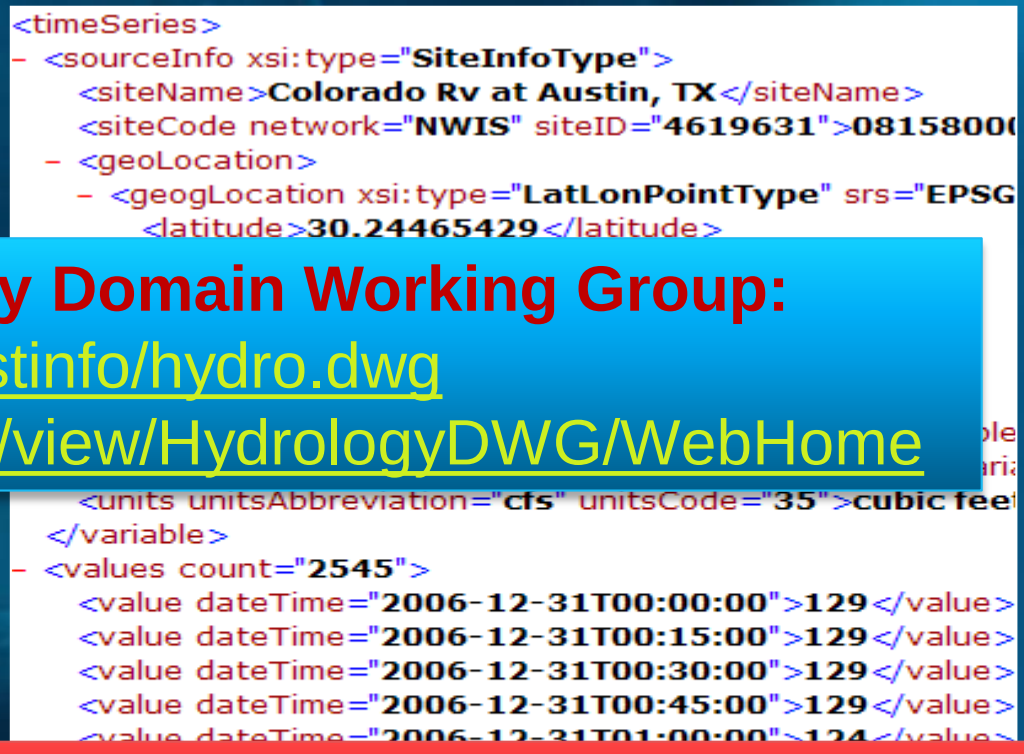
Desktop clients

- ArcGIS
- Matlab
- IDL, R
- MapWindow
- Excel
- Programming (Java, C, VB,…)
- Modeling (OpenMI)

Standard Water Data Services

Microsoft Research

- Set of query functions
- Returns data in **WaterML**



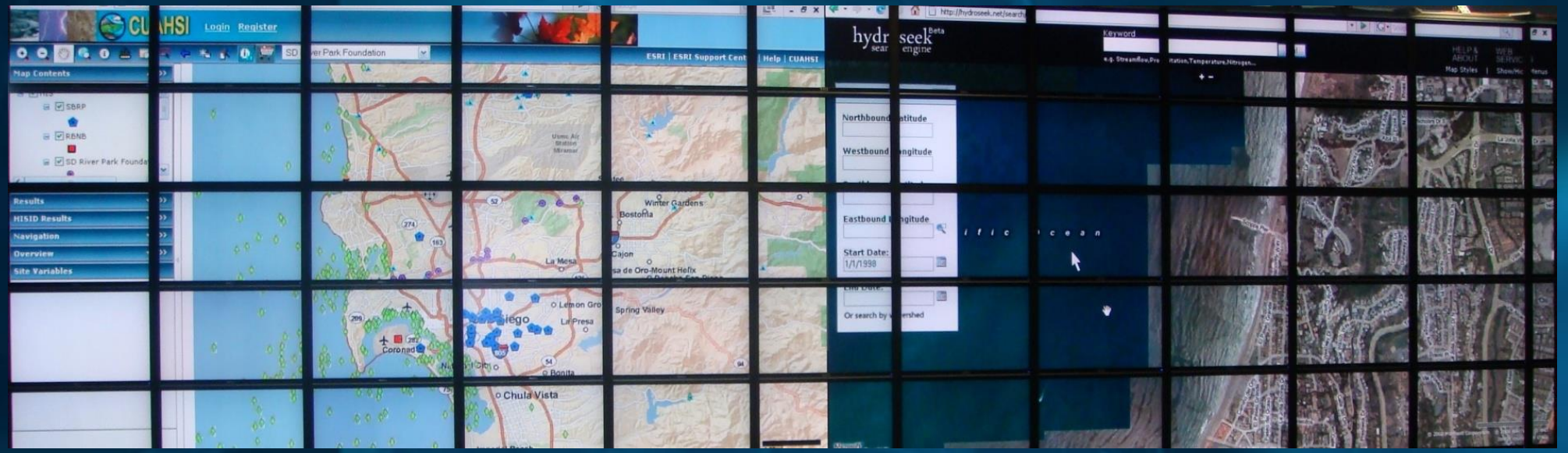
Next Step: WaterML 2.0; OGC Hydrology Domain Working Group:
https://lists.opengeospatial.org/mailman/listinfo/hydro.dwg
http://external.opengis.org/twiki_public/bin/view/HydrologyDWG/WebHome

NWIS Daily Values (discharge), NWIS Ground Water, NWIS Unit Values (real time), NWIS Instantaneous Irregular Data, EPA STORET, NCDC ASOS, DAYMET, MODIS, NAM12K, USGS SNOTEL, ODM (multiple sites)

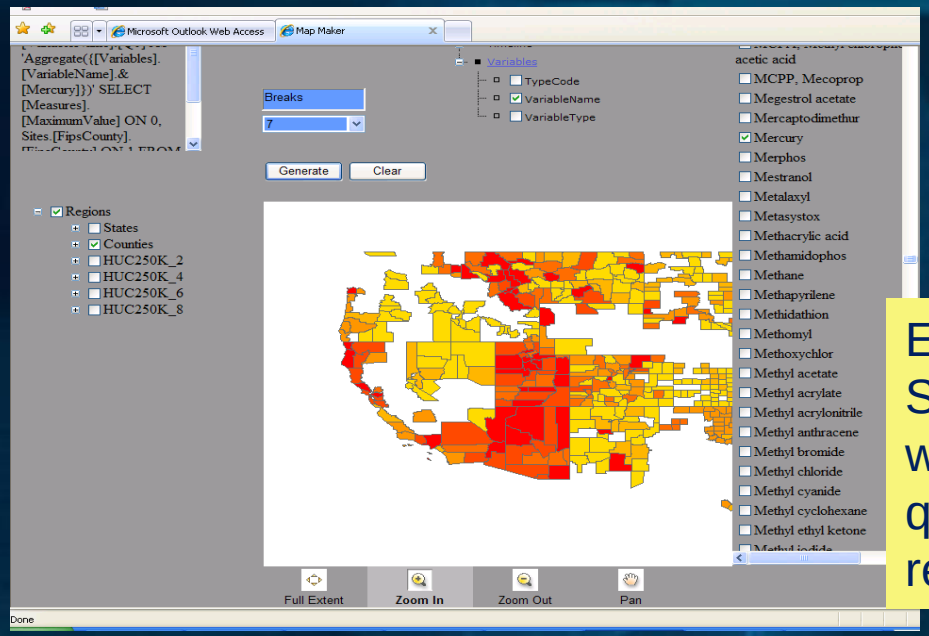
Visualization and Analysis of Large Datasets

Microsoft Research

- **Tiled wall**

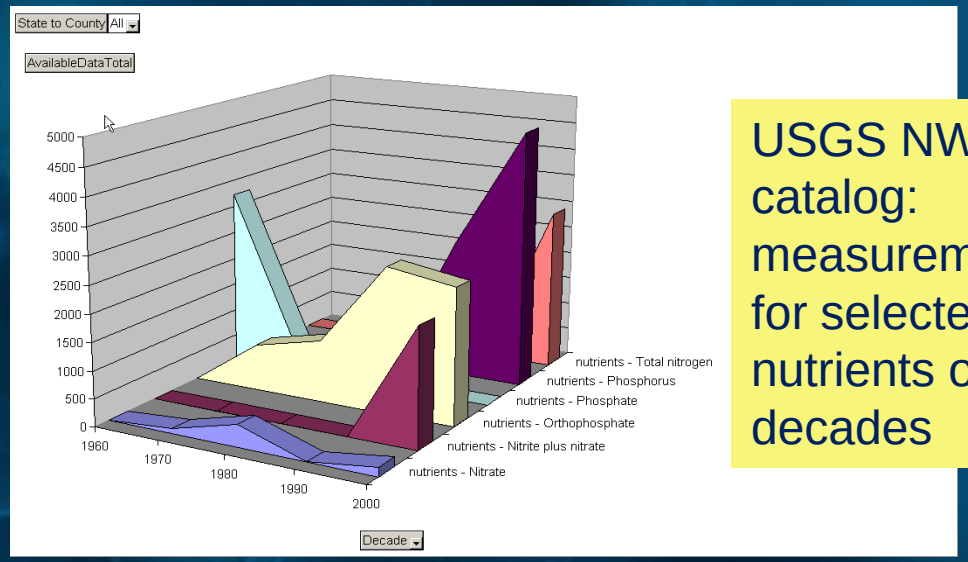


- **OLAP cubes for repositories**



EPA STORET water quality repository

- **OLAP cubes for catalogs**



USGS NWIS catalog: measurement totals for selected nutrients over decades

CUAHSI HIS Central - Windows Internet Explorer

http://hiscentral.cuahsi.org/pub_network.aspx?n=12

File Edit View Favorites Tools Help

Convert Select

CUAHSI HIS Central

Page Tools

CUAHSI HIS

Sharing hydrologic data

Login Register

Home All Data Services

Mud Lake Utah

Utah Water Research Laboratory, Utah State University

MudLake

UtahState UNIVERSITY

http://his02.usu.edu/MudLake/cuahsi_1_0.asmx?WSDL

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

Service Statistics:

Sites: 5
Variables: 26
Values: 792406

Geographic Extent:
42.25035
-111.3405 -111.2884
42.12042

Abstract

As part of a Utah State University Water Initiative funded research project we are working together with the U.S. Fish and Wildlife Service and the State of Idaho Department of Environmental Quality to collect continuous water quality monitoring data to investigate the sediment and nutrient budget of Mud Lake National Wildlife Refuge, Idaho.

Citation

Data collected in Mud Lake at the Bear Lake National Wildlife Refuge by Jeff Horsburgh, David Stevens, Nancy Mesner, and Cody Allen of Utah State University

Done Internet 100%

HIS Central Metadata Catalog and Service Registry

Managing Varying Semantics

In measurement units…

acre feet	acre-feet
micrograms per kilogram	micrograms per kilgram
FTU	NTU
mho	Siemens
ppm	mg/kg

In parameter names…

Nitrogen: e.g. NWIS parameter # 625 is **labeled 'ammonia + organic nitrogen',** Kjeldahl method is used for determination but not mentioned in parameter description. In STORET this parameter is referred to as Kjeldahl Nitrogen.

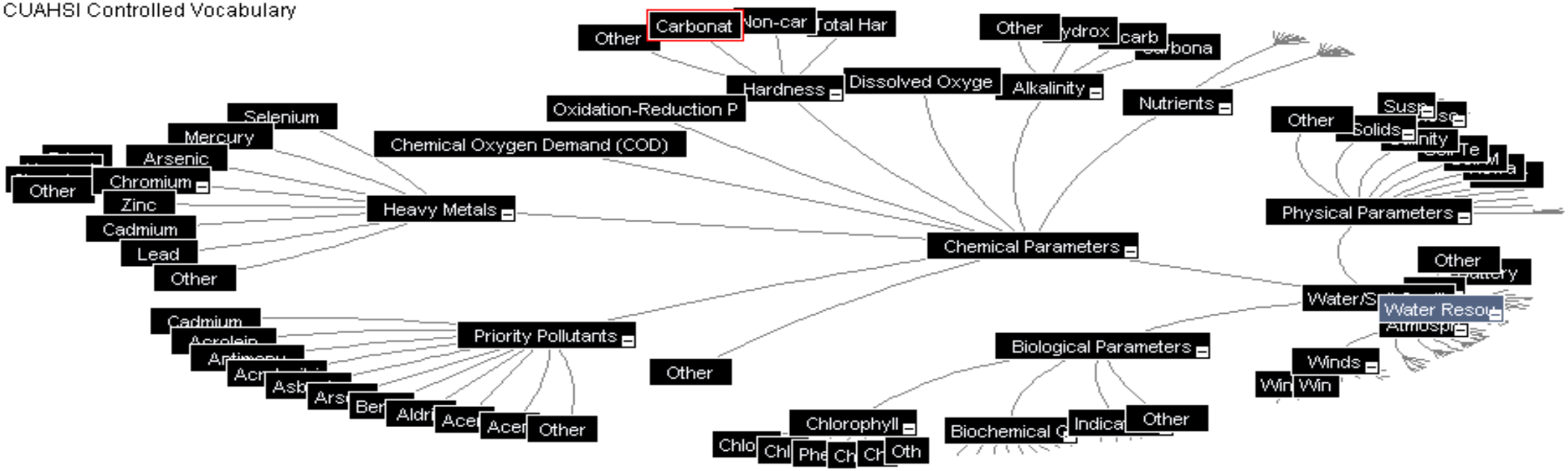
And: Dissolved oxygen

CUAHSI Data Annotation Tool - Windows Internet Explorer

http://water.sdsc.edu/tagger/tagger-demo.aspx

Semantic Tagging of Harvested Variables

CUAHSI Controlled Vocabulary



Star Tree™ created with Inxight VizServer™

inxight

Variable Name	Medium	
radiation, incoming par	air	select
	air	select
ge	surface water	select
	surface water	select
	surface water	select

Variable:

Mapping:

Search:

Search

	ConceptCode	Variable Name
delete	cadmium	transparency, water...
delete	subsurfhydro_other	orthphosphatephosph...
delete	turbidity	turbidity
delete	windgustvelocityabs...	temperature
delete	batteryvoltage	transparency, water...

HIS Central Semantic Tagging Tool

More Issues to Consider:

Microsoft Research

- Preservation of hydrologic data:
  - What is the time horizon for preservation; what are archival units and metadata; how are data relationships and transformations maintained; how are archives migrated
- Model integration:
  - Model coupling frameworks; managing and re-using model output; integration of observations and model results across scales; maintaining provenance and semantics in workflows and coupling; explanation facilities
- Managing community semantics process:
  - Community lexicon; a system of formal ontologies; semantic resource tagging; mechanism for evolving ontologies
- Your turn!