The CUAHSI Community Hydrologic Information System

David Tarboton, David Maidment, Ilya Zaslavsky, Dan Ames, Jon Goodall, Jeffery Horsburgh, Kim Schreuders

http://his.cuahsi.org/
What is CUAHSI?

Consortium of Universities for the Advancement of Hydrologic Science, Inc.

- 110 US University members
- 6 affiliate members
- 12 International affiliate members (as of March 2009)

Infrastructure and services for the advancement of hydrologic science and education in the U.S.

http://www.cuahsi.org/
The CUAHSI Hydrologic Information System (HIS) is an internet based system to support the sharing of hydrologic data. It is comprised of hydrologic databases and servers connected through web services as well as software for data publication, discovery and access.
Web Paradigm

Catalog (Google)

Web Server (CNN.com)  Search

Catalog harvest  Access

Browser (Firefox)
CUAHSI Hydrologic Information System
Services-Oriented Architecture

HydroCatalog
Data Discovery and Integration

HydroServer
Data Publication
ODM, Geo Data

HydroDesktop
Data Analysis and Synthesis
OpenMI, R

Metadata Services
Search Services

WaterML, Other OGC Standards

Information Model and Community Support Infrastructure
Let’s see some of it

- http://icewater.usu.edu/
- http://hydroservlet.codeplex.com
- http://hydrodesktop.codeplex.com
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
  - GetSites
  - GetSiteInfo
  - GetVariableInfo
  - GetValues

- Returns data in **WaterML**

```
<timeSeries>
- <sourceInfo xsi:type="SiteInfoType">
  <siteName>Colorado River at Austin, TX</siteName>
  <siteCode network="NWIS" siteID="4619631">08158001</siteCode>
  <geoLocation>
    <geogLocation xsi:type="LatLonPointType" src="EPSG">
      <latitude>30.24465429</latitude>
      <longitude>-97.694448</longitude>
    </geoLocation>
  </geoLocation>
</sourceInfo>
- <variable>
  <variableCode vocabulary="NWIS" default="true" variable="Discharge, cubic feet per second">
    <units unitsCode="35">cubic feet</units>
    <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="2006-12-31T1:00:00">124</value>
    <value dateTime="2006-12-31T1:15:00">129</value>
    <value dateTime="2006-12-31T1:30:00">124</value>
    <value dateTime="2006-12-31T1:45:00">124</value>
  </variable>
</timeSeries>
```
WaterML as a Web Language

USGS Streamflow data in WaterML language

Discharge of the San Marcos River at Luling, TX
June 28 - July 18, 2002

This is the WaterML GetValues response from NWIS Daily Values
Jointly with World Meteorological Organization

"Welcome to the Hydrology Domain Working Group"

**Work Plan**

**Meetings**
- Atlanta OGC TC Meeting - 17 September 2008
- Valencia OGC TC Meeting - 4 December 2008
- Athens OGC TC Meeting - 30 March 2009
- Boston OGC TC Meeting - 22 June 2009
- Darmstadt OGC TC Meeting - 29 September 2009
- Mountain View OGC TC Meeting - 8 December 2009
- Ispra Hydrology DWG Workshop - 15-18 March 2010 - Agenda
- Silver Spring OGC TC Meeting - 15 June 2010
- Toulouse OGC TC Meeting - 22 September 2010
- Toulouse Hydrology DWG Workshop - 21-22 September 2010 - Agenda
- Sydney Hydrology DWG Meeting - 1 December 2010

**Interoperability Experiments**
- GroundwaterInteroperabilityExperiment
- SurfacewaterInteroperabilityExperiment

**Evolving WaterML into an International Standard**

Meets every 3 months

Teleconferences most weeks

WaterML Version 2 standard being proposed

Vote for adoption 3-6 months later

To be open for public comment April to May 2011

http://www.opengeospatial.org/projects/groups/waterml2.0swg
HydroDesktop
Desktop Hydrologic Information System

Observations

Weather and Climate

Modeling

GIS

Remote Sensing
HydroDeskop

- Add shapefiles to map
- Change symbology and labels
- Print and export map
- GIS toolbox

Hydrology
- Search for data
- Download data
- Display time series
- Export data

http://hydrodesktop.codeplex.com
12800 total downloads
2000 code commits
25 registered developers
HydroModeler

An integrated modeling environment based on the Open Modeling Interface (OpenMI) standard and embedded within HydroDesktop

In development at the University of South Carolina by Jon Goodall, Tony Castronova, Mehmet Ercan, Mostafa Elag, and Shirani Fuller
Integration with “R” Statistics Package
http://hydroserver.codeplex.com

- A platform for publishing space-time hydrologic datasets that:
  - Autonomous with local control of data
  - Part of a distributed system that makes data universally available
- Basis for Experimental Watershed or Observatory data management system
- Standards based approach to data publication
  - Accepted and emerging standards for data storage and transfer (OGC, WaterML)
- Built on established software
  - MS SQL Server, ArcGIS server
- Open Source Community Code Repository
  - Sustainability
Ongoing Data Collection

Point Observations Data

Historical Data Files

GIS Data

Internet Applications

GetSites
GetSiteInfo
GetVariableInfo
GetValues

WaterML
WaterOneFlow
Web Service

ArcGIS Server

Data presentation, visualization, and analysis through Internet enabled applications
Observation Data Model for hydrologic and environmental measurements

The way that data is organized can enhance or inhibit the analysis that can be done.
Why an Observations Data Model

• Provides a common persistence model for observations data

• Syntactic heterogeneity (File types and formats)
• Semantic heterogeneity
  – Language for observation attributes (structural)
  – Language to encode observation attribute values (contextual)

• Publishing and sharing research data
• Metadata to facilitate unambiguous interpretation
• Enhance analysis capability
Scope

• Focus on Hydrologic Observations made at a point
• Exclude Remote sensing or grid data.
• Primarily store raw observations and simple derived information to get data into its most usable form.
• Limit inclusion of extensively synthesized information and model outputs at this stage.
What are the basic attributes to be associated with each single data value and how can these best be organized?

<table>
<thead>
<tr>
<th>Value</th>
<th>Offset</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>DateTime</td>
<td>OffsetType/Reference Point</td>
<td>Quality Control</td>
</tr>
<tr>
<td>Variable</td>
<td>Source/Organization</td>
<td>Level</td>
</tr>
<tr>
<td>Location</td>
<td>Censoring</td>
<td>Sample Medium</td>
</tr>
<tr>
<td>Units</td>
<td>Data Qualifying Comments</td>
<td>Value Type</td>
</tr>
<tr>
<td>Interval</td>
<td></td>
<td>Data Type</td>
</tr>
<tr>
<td>(support)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CUAHSI Observations Data Model

- A relational database at the single observation level (atomic model)
- Stores observation data made at points
- Metadata for unambiguous interpretation
- Traceable heritage from raw measurements to usable information
- Standard format for data sharing
- Cross dimension retrieval and analysis

A data value $v_i(s,t)$ in space, $s$, time, $t$, and variables, $V_i$.
Data Storage – Relational Database

### Values

<table>
<thead>
<tr>
<th>Value</th>
<th>Date</th>
<th>Site</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.5</td>
<td>3/3/2007</td>
<td>1</td>
<td>Streamflow</td>
</tr>
<tr>
<td>4.2</td>
<td>3/4/2007</td>
<td>1</td>
<td>Streamflow</td>
</tr>
<tr>
<td>33</td>
<td>3/3/2007</td>
<td>2</td>
<td>Temperature</td>
</tr>
<tr>
<td>34</td>
<td>3/4/2007</td>
<td>2</td>
<td>Temperature</td>
</tr>
</tbody>
</table>

### Sites

<table>
<thead>
<tr>
<th>Site</th>
<th>Name</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cane Creek</td>
<td>41.1</td>
<td>-103.2</td>
</tr>
<tr>
<td>2</td>
<td>Town Lake</td>
<td>40.3</td>
<td>-103.3</td>
</tr>
</tbody>
</table>
Why Use a RDBMS

- Mature and stable technology
- Structured Query Language (SQL)
- Sharing of data among multiple applications
  - Data integrity and security
  - Access by multiple users at the same time
  - Tools for backup and recovery
- Reduced application development time
Discharge, Stage, Concentration and Daily Average Example
Independent of, but can be coupled to Geographic Representation

**ODM**
- Observations Data Model
  - Sites
    - SiteID
    - SiteCode
    - SiteName
    - Latitude
    - Longitude
    - ...
  - CouplingTable
    - SiteID
    - HydroID

**e.g. Arc Hydro**
- Feature
  - HydroPoint
    - HydroID
    - HydroCode
    - FType
    - Name
    - JunctionID
  - Waterbody
    - HydroID
    - HydroCode
    - FType
    - Name
    - AreaSqKm
    - JunctionID
  - Watershed
    - HydroID
    - HydroCode
    - DrainID
    - AreaSqKm
    - JunctionID
    - NextDownID
  - ComplexEdgeFeature
    - HydroID
    - HydroCode
    - ReachCode
    - Name
    - LengthKm
    - LengthDown
    - FlowDir
    - FType
    - EdgeType
    - Enabled
  - SimpleJunctionFeature
    - HydroID
    - HydroCode
    - NextDownID
    - LengthDown
    - DrainArea
    - FType
    - Enabled
    - AncillaryRole
  - HydroNetwork
    - HydroID
    - HydroCode
    - NextDownID
    - LengthDown
    - DrainArea
    - FType
    - Enabled
    - AncillaryRole
  - Flowline
  - Shoreline

(Independent of, but can be coupled to Geographic Representation)
Stage and Streamflow Example
Loading data into ODM

- Interactive OD Data Loader (OD Loader)
  - Loads data from spreadsheets and comma separated tables in simple format
- Scheduled Data Loader (SDL)
  - Loads data from datalogger files on a prescribed schedule.
  - Interactive configuration
- SQL Server Integration Services (SSIS)
  - Microsoft application accompanying SQL Server useful for programming complex loading or data management functions
Managing Data Within ODM ODM Tools

- **Query and export** – export data series and metadata
- **Visualize** – plot and summarize data series
- **Edit** – delete, modify, adjust, interpolate, average, etc.
Publication of Spatial (GIS) Datasets

• Publishing spatial datasets using ArcGIS Server
  – Using OGC standards that can be consumed by a number of GIS clients
  – WMS, WFS, WCS
Data Presentation Via a Map Interface

- Internet Map Server built using ArcGIS
- Web browser client
- Combine spatial data and observational data
- Launch data visualization tools
- Based on a “Region”

http://icewater.usu.edu/map/
Data Preview, Visualization, and Analysis
Time Series Analyst

- Web Browser Client
- Multiple ODM Database Support
- Variety of plot types
- Descriptive statistics
- Linked to the map application
- Data preview and download

http://icewater.usu.edu/tsa/
HydroServer Capabilities Web Service

- Publish capabilities of each HydroServer
  - Listing of published observational data services
  - Listing of published spatial data services
- Supports automatic cataloging of available services at HIS Central
- Makes HydroServers self-describing

![HISNodeCapabilities Web Service](image)
Overcoming Semantic Heterogeneity

• ODM Controlled Vocabulary System
  – ODM CV central database
  – Online submission and editing of CV terms
  – Web services for broadcasting CVs

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Investigator 1: “Temperature, water”</td>
<td>...</td>
</tr>
<tr>
<td>Investigator 2: “Water Temperature”</td>
<td>Sunshine duration</td>
</tr>
<tr>
<td>Investigator 3: “Temperature”</td>
<td>Temperature</td>
</tr>
<tr>
<td>Investigator 4: “Temp.”</td>
<td>Turbidity</td>
</tr>
<tr>
<td></td>
<td>...</td>
</tr>
</tbody>
</table>
Dynamic controlled vocabulary moderation system

ODM Data Manager

ODM Tools

Local ODM Database

Local Server

ODM Website

ODM Controlled Vocabulary Moderator

XML

ODM Controlled Vocabulary Web Services

Master ODM Controlled Vocabulary

http://his.cuahsi.org/mastercvreg.html

From Jeff Horsburgh
37 Water Data Services on HIS Central from 12 Universities

- University of Maryland, Baltimore County
- Montana State University
- University of Texas at Austin
- University of Iowa
- Utah State University
- University of Florida
- University of New Mexico
- University of Idaho
- Boise State University
- University of Texas at Arlington
- University of California, San Diego
- Idaho State University

Dry Creek Experimental Watershed (DCEW)
(28 km² semi-arid steep topography, Boise Front)

68 Sites
24 Variables
4,700,000+ values

Published by Jim McNamara, Boise State University
Water Agencies and Industry

- USGS, NCDC, Corps of Engineers publishing data using HIS WaterML
- OGC Hydrology Domain Working Group evaluating WaterML as OGC standard
- ESRI using CUAHSI model in ArcGIS.com GIS data collaboration portal
- Kisters WISKI support for WaterML data publication
- Australian Water Resources Information System Water Accounting System has adopted aspects of HIS
- NWS West Gulf River Forecast Center Multi-sensor Precipitation Estimate published from ODM using WaterML
# Federal Agency Water Data Services at HISCentral (10/2010)

<table>
<thead>
<tr>
<th>Network Name</th>
<th>Site Count</th>
<th>Value Count</th>
<th>Earliest Observation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NWISDV</td>
<td>31,800</td>
<td>304,000,000</td>
<td>01/01/1861</td>
<td>WaterML-compliant GetValues service from NWIS, catalog ingested</td>
</tr>
<tr>
<td>EPA</td>
<td>236,000</td>
<td>78,000,000</td>
<td>01/11/1900</td>
<td>SOAP wrapper over WQX services, catalog ingested</td>
</tr>
<tr>
<td>NWISUV</td>
<td>11,800</td>
<td>84,500,000</td>
<td>60 DAYS</td>
<td>WaterML-compliant GetValues Service, catalog ingested</td>
</tr>
<tr>
<td>NCDC ISH</td>
<td>11,600</td>
<td>3,000,000*</td>
<td>1/1/2005</td>
<td>WaterML-compliant GetValues service from NCDC</td>
</tr>
<tr>
<td>NCDC ISD</td>
<td>24,800</td>
<td>18,200,000</td>
<td>1/1/1892</td>
<td>WaterML-compliant GetValues service from NCDC</td>
</tr>
<tr>
<td>NWISIID</td>
<td>376,000</td>
<td>86500,000</td>
<td>9/1/1867</td>
<td>SOAP wrapper over NWIS web site, catalog ingested</td>
</tr>
<tr>
<td>NWISGW</td>
<td>834,000</td>
<td>8,490,000</td>
<td>1/1/1800</td>
<td>SOAP wrapper over NWIS web site, catalog ingested</td>
</tr>
<tr>
<td>RIVERGAGES</td>
<td>1,300</td>
<td>264,000,000</td>
<td>1/1/2000</td>
<td>WaterML compliant REST services from Army Corps of Engineers</td>
</tr>
</tbody>
</table>

* Estimated
USGS Unit Values Data

Real time streamflow data over the last 60 days

11188 sites, nationally for the US

Published by USGS National Water Information System
HISCentral Content (11/2010)

Map integrating NWIS, STORET, & Climatic Sites

58 public services
18,000+ variables
1.96+ million sites
23.3 million series
Referencing 5.1 billion data values

Available via HISCentral discovery services
Available via GetValues requests
Summary

• **Data Storage** in an *Observations Data Model (ODM)* and publication through *HydroServer*

• **Data Access** through internet-based *Water Data Services* using a consistent data language, called *WaterML* from *HydroDesktop*

• **Data Discovery** through a *National Water Metadata Catalog* and thematic keyword search system at *HIS Central*

• **Integrated Modeling and Analysis** within *HydroDesktop*

The combination of these capabilities creates a common window on water observations data for the United States unlike any that has existed before.
• Learn about the CUAHSI-HIS System
• Share your work with information systems and large scale datasets
• Share your use of hydrologic data for teaching
• Interact with other users
• Share your work linking data and modeling
• Show science enabled by HIS
• Hands-on workshops
• Contribute to the future of HIS

For information on presenting or attending see:
http://his.cuahsi.org/conference2011
Contact: David.Tarboton@usu.edu
Thanks! HIS Project Team and Sponsors

- **University of Texas at Austin** – David Maidment, Tim Whiteaker, James Seppi, Fernando Salas, Jingqi Dong, Harish Sangireddy
- **San Diego Supercomputer Center** – Ilya Zaslavsky, David Valentine, Tom Whitenack, Matt Rodriguez
- **Utah State University** – Jeff Horsburgh, Kim Schreuders, Stephanie Reeder, Edward Wai Tsui, Ravichand Vegiraju, Ketan Patil
- **University of South Carolina** – Jon Goodall, Anthony Castronova
- **Idaho State University** – Dan Ames, Ted Dunsford, Jiří Kadlec, Yang Cao, Dinesh Grover
- **Drexel University/CUNY** – Michael Piasecki
- **WATERS Network** – Testbed Data Managers
- **CUAHSI Program Office** – Rick Hooper, Yoori Choi, Conrad Matiuk
- **ESRI** – Dean Djokic, Zichuan Ye

[CUAHSI HIS](http://his.cuahsi.org/)