Advancing an Information Model for Environmental Observations

Jeffery S. Horsburgh

Anthony Aufdenkampe, Richard P. Hooper, Kerstin Lehnert, Kim Schreuders, David G. Tarboton
The Original Problem
Combining Similar Data from Disparate Sources

Slide from Michael Piasecki
CUAHSI Hydrologic Information System
A Services Oriented Architecture for Sharing Hydrologic Observations

HydroServer
Stores, organizes, and publishes data and metadata

Metadata Services

Data Services

HIS Central
Harvests metadata from the data servers and allows efficient data search by clients

Search Services

HydroDesktop
Provides users a convenient way to search, map, graph, and analyze data

Slide from Steven Brown
A **data source** operates an observation network

A **network** is a set of observation sites

A **site** is a point location where one or more variables are measured

A **variable** is a property describing the flow or quality of water

A **value** is an observation of a variable at a particular time

A **qualifier** is a symbol that provides additional information about the value

---

**USGS NWIS**

**Data Source**

**NWIS Sites**

**Network**

**Sites**

**Variables**

**Observation**

{Value, Time, Qualifier}

**San Marcos River at Luling, TX**

**Discharge**

18,700 cfs, 3 July 2002
Data Values – Indexed by “What-Where-When”

- Space, $S$
- Time, $T$
- Variables, $V$

$v_i(s,t)$

“Where”

“When”

“What”

A data value
There are $n$ measurements of Variable $V_i$ at Site $S_j$ from time $t_1$ to time $t_2$. 
Observations Data Model (ODM)

- A relational database at the single observation level
- Metadata for unambiguous interpretation
- Traceable heritage from raw measurements to usable information
- Promote syntactic and semantic consistency

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">08158001</siteCode>
  - <geoLocation>
    - <geoLocation xsi:type="LatLonPointType" srs="EPSG">
      - <latitude>30.24465429</latitude>
      - <longitude>-97.694448</longitude>
    </geoLocation>
  </geoLocation>
</sourceInfo>
```

```
<variable>
  - <variableCode vocabulary="NWIS" default="true">Discharge</variableCode>
  - <variableName>Discharge, cubic feet per second</variableName>
  - <units unitsAbbreviation="cfs" unitsCode="35">cubic feet per second</units>
</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="2006-12-31T1:00:00">124</value>
  <value dateTime="2006-12-31T1:15:00">124</value>
  <value dateTime="2006-12-31T1:30:00">124</value>
  <value dateTime="2006-12-31T1:45:00">124</value>
</values>
```
HydroCatalog

http://hiscentral.cuahsi.org

- Service registry
- Metadata harvester
- Catalog database
- Semantic tagging application
- Data discovery services
HydroDesktop – Data Access and Analysis

Thematic keyword search

Integration from multiple sources

Search on space and time domain
A Common Information Model Has Enabled

• A greater degree of semantic and syntactic homogeneity across data sources
• Ability to catalog and provide semantically enabled search services across multiple disparate data sources
• Evolution toward community standards for sharing hydrologic data
Limitations

• Works great for a limited class of hydrologic observations (e.g., point time series), BUT...
  – Does not adequately support some types of \textit{ex situ} observations
  – Does not adequately support observations on geometries other than points
  – Does not adequately describe the “feature of interest” or the “sampled feature” or “Site Type”
  – Does not provide adequate ability to record provenance and annotate observations
  – ...

Both *in situ* and *ex situ* data are critical for analysis of the critical zone.
Coupled Human/Natural Systems

Snow depth, distribution, density

Precipitation

Reservoir inflows, storage, releases

Irrigation, Evapotranspiration

Streamflow quantity, quality

Data representing multiple hydrologic features with complex geometries.
A More Flexible Information Model for Observations

- A number of important advancements have emerged
  - **Information Model**
    - Open Geospatial Consortium Observations & Measurements
  - **Service Interfaces**
    - Open Geospatial Consortium Standards – particularly Sensor Observation Services (SOS)
  - **Semantics and Integration**
    - Scientific Observations Network (SONet)
  - **Data Storage Model**
    - Microsoft Research and SDSC Environmental Data Model (EDM)
International Standardization of WaterML

Hydrology Domain Working Group
- working on WaterML 2.0
- organizing Interoperability Experiments focused on different sub-domains of water
- towards an agreed upon feature model, observation model, semantics and service stack

Timeline

- Groundwater IE
  - GSC+USGS
  - Dec 09 – Dec 10
- Surface Water IE
  - CSIRO+many
  - Jun 10 – Sep 11
- Forecasting IE
  - NWS+Deltares?
  - Sep 11 – Sep 12?
- Water Quality IE
- Water Use IE

Iterative Development

- Feedback to standards and services
- Conclude IE Demonstration
- Develop IE

Slide from Ilya Zaslavsky
ODM 2.0

• Approach: a core observational data model with flexible extensions
  – \textit{Samples extension} – better handle storage of sample information and observations derived from \textit{ex situ} analyses
  – \textit{Field sensor extension} – better handle storage of sensor deployment information and \textit{in situ} observations
  – \textit{Provenance and annotation} – capturing more of the context of observations
• A more robust “feature model” to better describe the geographic context of observations
• Enhanced semantics
• Harmonization with WaterML
Thank you!