Using CUAHSI HIS to Support Large Scale Collaborative Research in Utah

Jeffery S. Horsburgh

Amber Jones, Stephanie Reeder, James Patton, Juan Caraballo, Mauriel Ramirez, Nate Mouzon

CUAHSI HIS Cyberseminar
5-1-2013
innovative Urban Transitions and Aridregion Hydro-sustainability (iUTAH)

- Statewide effort
- $20 million competitive award from NSF EPSCoR
- Research capacity building
- Interdisciplinary and multi-institution
- Focused on sustainable management of Utah’s water resources
iUTAH Research Focus Areas

1. **Eco-hydrology**
   - Expand Utah’s capacity in the natural sciences through instrumentation of 3 watersheds

2. **Social and Engineered Water Systems**
   - Studying demographic characteristics, water use behaviors, water infrastructure, and other measures of urban form

3. **Interdisciplinary Modeling and Visualization**
   - Development of interdisciplinary models of socio-eco-hydrological systems to determine how changes in water availability and use alter water quantity and quality
iUTAH Modeling and Data Federation

**HPC@USU**
- Primary data organization and storage
- Data sharing, publication, and archival
- Data integration and synthesis
- Integrated modeling and CI Support
- HPC support

**UWyoming (EPSCoR CI Track 2) and National HPC Centers**
- Redundant data storage
- HPC support
- Modeling and simulation
- Collaboration technologies

**UU Center for High Performance Computing**
- Redundant data storage
- HPC support
- Modeling and simulation
- Collaboration technologies

**Fulton Supercomputing Lab**
- BYU

**Agency Data**
- USGS NWIS
- NRCS SNOTEL
- Utah DWQ
- Etc.

**Primary Data Storage**
- CUAHSI HIS
- DataONE

**National Science Data Networks**
- CUAHSI HIS
- DataONE

**iUtah Field Sites**
- Little Bear River
- Red Butte Creek
- Provo River

**Utah Field Station Network**

**UVU**

**UEN and K-12 Sites**

**Intermountain Network Node**

**Research@UEN Network**

**UEN Network**

**BYU**

**Intermountain Network Node**
iUTAH MDF – Information Flow

- Field sites and other data (e.g., agencies)
- iUTAH Modeling and Data Federation
- Modeling, visualization, and analysis
iUTAH MDF Focus Areas

1. Data management support for data from iUTAH facilities
   – GAMUT – Gradients Along Mountain to Urban Transitions
   – GIRF – Green Infrastructure Research Facility

2. Support for data discovery and access
   – iUTAH facility and investigator data
   – Agency data
   – National data networks

3. Support for integrated modeling

4. Data/model/resource sharing and collaboration
How can we create a hardware platform that supports the diverse cyberinfrastructure needs of iUTAH?
iUTAH CI Development Approach

- Where possible, leverage and adopt existing cyberinfrastructure components

- Collaborate with other CI development activities to get needed functionality

- Develop pieces that we need
Required Functionality

• Research platform
  – Development, prototyping, and testing of servers, software applications, and services on multiple platforms (e.g., Windows, Linux)
  – Host machines allocated for modeling, analysis, and computational tasks

• Production data service and application hosting platform
  – Web servers
    • Data web services
    • Web applications
  – Database servers
  – Map servers
  – File servers
  – Data harvesters and automated data processing applications
Hardware

Required servers and software stacks implemented on virtual machines

Many virtual machines can be hosted on a single virtual host server

Tiered storage
Tier 1 and 2: Databases and operating systems
Tier 3: Lower demand file storage and access
Hardware Capabilities

• Quickly spin up virtual machines that implement different operating systems and platforms (e.g., Windows, Linux)

• Use shared computational and storage resources rather than requiring a physical machine for each purpose

• Create and manage both development and production servers using shared hardware

• Hot-swap virtual machines across physical host machines using virtualization software
  – Ensure failover for production virtual servers
  – Efficiently allocate resources to multiple machines
Initial Specifications

• 3 virtual host servers using shared storage arrays
  – Dual 8 core processors
  – 128 GB RAM

• 7.2 TB high performance shared storage array
  – Virtual machine operating systems
  – Relational databases

• 72 TB high capacity shared storage array
  – File storage
  – Archival

• Integrated with USU Central IT VMWare infrastructure and enterprise data center

Available storage will be expanded in subsequent years
The Larger Utah EPSCoR Cyberinfrastructure Picture

USU VMWare Cluster (iUTAH, CI-WATER)

Virtual machine backup
Dataset backup

USU HPC (existing)

Simulation data
Post processing data

UofU VMWare Cluster (CI-WATER, iUTAH)

Virtual machine backup
Dataset backup

UofU Large Scale Storage (CI-WATER, iUTAH)

Simulation data
Post processing data

UofU CHPC (existing)

U. Wyoming Mt. Moran (CI-WATER)

Simulation data
Post processing data

Fulton Supercomputing Lab (existing)

BYU Brigham Young University
What is the design of a cyberinfrastructure that enables standardized data collection and management for a network of aquatic and terrestrial monitoring sites managed by a consortium of disparate organizations?
Gradients Along Mountain to Urban Transitions (GAMUT) Network

- Sensor network developed between USU/UofU/BYU
- Mix of aquatic and terrestrial in situ and re-locatable sensors
- Measure aspects of water inputs and outputs and water quality over gradient
- Deployed in three watersheds
Gradients Along Mountain to Urban Transitions (GAMUT) Network

- 3 watersheds have similar water source (high elevation snow) but different land use transitions
- Logan River: irrigated agriculture transitioning to moderate density urban at moderate pace
- Red Butte Creek: highly urbanized
- Provo River: irrigated agriculture transitioning to low density urban at rapid pace
Gradients Along Mountain to Urban Transitions (GAMUT) Network

Table 1. Parameters to be measured by the iUTAH Climate and Ecohydrology Sensor Network.

<table>
<thead>
<tr>
<th>Sensor Type</th>
<th>Fundamental Suite</th>
<th>Enhanced/Urban Suite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Terrestrial Sensors</td>
<td>Barometric pressure</td>
<td>Barometric pressure</td>
</tr>
<tr>
<td></td>
<td>Wind speed and direction</td>
<td>Wind speed and direction</td>
</tr>
<tr>
<td></td>
<td>Air temperature</td>
<td>Air temperature</td>
</tr>
<tr>
<td></td>
<td>Relative humidity</td>
<td>Relative humidity</td>
</tr>
<tr>
<td></td>
<td>Precipitation</td>
<td>Precipitation</td>
</tr>
<tr>
<td></td>
<td>Snow depth</td>
<td>Snow depth</td>
</tr>
<tr>
<td></td>
<td>Soil temperature, moisture, conductivity</td>
<td>Soil temperature, moisture, conductivity</td>
</tr>
<tr>
<td></td>
<td>Solar radiation (net radiation and PAR)</td>
<td>Solar radiation (net radiation and PAR)</td>
</tr>
<tr>
<td>Aquatic Sensors</td>
<td>Stream stage</td>
<td>Stream stage</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td>Temperature</td>
</tr>
<tr>
<td></td>
<td>Electrical Conductivity</td>
<td>Electrical Conductivity</td>
</tr>
<tr>
<td></td>
<td>pH</td>
<td>pH</td>
</tr>
<tr>
<td></td>
<td>Dissolved oxygen</td>
<td>Dissolved oxygen</td>
</tr>
<tr>
<td></td>
<td>Turbidity</td>
<td>Turbidity</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total algae (chlorophyll a + phycocyanin)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>fDOM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Nitrate</td>
</tr>
</tbody>
</table>
Challenges to Managing Sensor Data

- Volume of data
- Data heterogeneity
- Multiple watersheds
- Multiple institutions
- Scale
- Synchronize timing, data access, equipment tracking
- Standardize data QA/QC

Rainfall and Meteorology

Water quality

GIS

Soil water

Water quantity

Groundwater
HydroServer
CUAHSI Open Source Hydrologic Data Tools

Observations Data Model

ODM Data Loader

ODM Streaming Data Loader

ODM Tools

WaterOneFlow and WaterML
Data Loading and Storage

iUTAH Modeling and Data Federation Facility

Primary Databases at USU

Backup Databases at UofU
Data Loading and Storage

- Observations Data Model (ODM): relational database at the single observation level
- Metadata for unambiguous interpretation
- Traceable heritage from raw measurements to usable information
- Promote syntactic and semantic consistency
- Cross dimension retrieval and analysis

Loading Data Streams
Data Visualization and Management

iUTAH Modeling and Data Federation Facility

ODM

Remote Data Managers

ODM Tools
ODM Tools Python
Data Visualization and Management

- **Plot Display Options**
- **Date Range Restrictions**
- **Dynamic Zooming and Panning**
- **Multiple Plot Types**
- **Time Series Selection**

![Graph showing time series data with various plot types and interactive features.](image-url)
ODM Tools Python
Multiple Plot Types
ODM Tools Python
Time Series Selection and Filtering

Build Query
Filters on Data Series
Export Data Series
ODM Tools Python
Dockable Windows
Sensor Data Quality Control

USU-LBR-Paradise - Little Bear River at McMurdy Hollow near Paradise, Utah

USU-LBR-SFLower - South Fork Little Bear River below Davenport Creek near Avon, Utah

Strange anomalies

Dead battery
ODM Tools Python
Sensor Data Quality Control

Data Editing Tools
Python Code Console
Dynamic Data Editing Display
Python Script Editor
Equipment Management

- Track physical infrastructure: sensors, data loggers, batteries, etc.
- Track events: deployments, calibrations, site visits, factory servicings, etc.
- Connects to ODM where streaming data is stored
- Web interface
# Equipment Details

<table>
<thead>
<tr>
<th>Equipment Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Equipment ID:</strong> 99999999999</td>
</tr>
<tr>
<td><strong>Serial Number:</strong> 99999999999</td>
</tr>
<tr>
<td><strong>Model Name:</strong> DTS-12</td>
</tr>
<tr>
<td><strong>Description:</strong> Forest Technology Systems DTS-12 Digital Turbidity Sensor</td>
</tr>
<tr>
<td><strong>Purchase Date:</strong> 1/20/2013</td>
</tr>
<tr>
<td><strong>Notes:</strong> Measures water turbidity and water temperature.</td>
</tr>
<tr>
<td><strong>Owner Institution:</strong> Utah State University</td>
</tr>
<tr>
<td><strong>Owner Contact:</strong> Michelle Baker</td>
</tr>
<tr>
<td><strong>Owner Address:</strong> 5305 Old Main Hill, Logan, UT 84322-5305</td>
</tr>
<tr>
<td><strong>Owner Phone:</strong> 1.435.797.7131</td>
</tr>
<tr>
<td><strong>Email:</strong> <a href="mailto:michelle.baker@usu.edu">michelle.baker@usu.edu</a></td>
</tr>
</tbody>
</table>

## Manufacturer

- **Vendor Name:** Forest Technology Systems
- **Contact Name:** Someguy Thatworksthere
- **Address:** 1123 Fir Avenue, Suite C., Blaine, WA, 98230
- **Phone:** 1.800.548.4264
- **Email:** sales@ftshydrology.com
- **Web Address:** [http://www.ftsenvironmental.com/](http://www.ftsenvironmental.com/)

## Vendor

- **Vendor Name:** Forest Technology Systems
- **Contact Name:** Someguy Thatworksthere
- **Address:** 1123 Fir Avenue, Suite C., Blaine, WA, 98230
- **Phone:** 1.800.548.4264
- **Email:** sales@ftshydrology.com
- **Web Address:** [http://www.ftsenvironmental.com/](http://www.ftsenvironmental.com/)
Site Visit Details | View full site visit details

Edit Description  Delete  View Full Site Visit History for this Site

Site Visit Location

Site Code: USU-LR-UWRL
Site Name: Logan River at the Utah Water Research Laboratory
Latitude: 43
Longitude: -111
Elevation: 4355 m

Site Visit Description

Crew: Jeff Horsburgh, Amber Jones
Begin Date Time: 2/22/2013 1:00 PM MST
End Date Time: 2/22/2013 2:00 PM MST
Environmental Observations: The weather was sunny. The river was low. Water was clear.
Site Visit Notes: None.

Field Activities Performed (click the activity type to view details)

<table>
<thead>
<tr>
<th>Activity Type</th>
<th>Begin Date/Time</th>
<th>End Date/Time</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td>2/22/2013 1:00 PM MST</td>
<td>2/22/2013 1:10 PM MST</td>
<td>Sensor Calibration</td>
<td>None</td>
</tr>
<tr>
<td>Deployment</td>
<td>2/22/2013 1:30 PM MST</td>
<td>2/22/2013 2:00 PM MST</td>
<td>Sensor Deployment</td>
<td>None</td>
</tr>
</tbody>
</table>

Copyright © 2013 iUTAH
This material is based upon work supported by the National Science Foundation (NSF) under Grant No. XXXXXXXX.
Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.
<table>
<thead>
<tr>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
<th>Sunday</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Display:**  ✔ Site Visits  ☐ Factory Service Events

Copyright © 2013 iUTAH

This material is based upon work supported by the National Science Foundation (NSF) under Grant No. XXXXXXXX.

Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.
ODM2

Observations Core

Samples Extension

Feature Model

Sensor Extension

Generic Extension

Common Semantics for Earth Observations

CUAHSI HIS

EarthChem

CZOData

IOOS

Domain Cyberinfrastructures
How can we enable and increase collaborative research and sharing of data and models through the innovative use of cyberinfrastructure?
iUTAH Modeling and Data Federation

data.iutahepscor.org

The iUTAH Modeling and Data Federation is an online system for sharing data, models, and other digital resources. Activities of the iUTAH CyberInfrastructure Team (CI Team) are focused on developing hardware and software systems and tools that improve iUTAH participants' capacity for data collection, organization, management, sharing, synthesis, to higher-level products, and integration with models.

Announcements

Data Inventory: As an initial step in developing the iUTAH Modeling and Data Federation, we are conducting a survey of existing and planned datasets. Click here to submit dataset information to the iUTAH data inventory or view the list of datasets that have been submitted.

Model Inventory: We are collecting information about models from a variety of disciplines that may be of interest to your iUTAH-related research. Click here to submit model information to the iUTAH data inventory or view the list of models that have been submitted.

Software Development

The beginnings of the iUTAH Modeling and Data Federation have been implemented at http://data.iutahepscor.org. This website will continue to be fleshed out as we develop new functionality of the next couple of years.

Data and Model Inventories

The iUTAH research teams will be developing new datasets and models to support iUTAH's research goals. However, there will also be extensive reuse of existing data resources and models from many different sources. The CI Team is developing tools that will help facilitate the process of identifying existing datasets and models that may be useful to the iUTAH efforts. For example, we are compiling a database of metadata describing existing and planned data resources. iUTAH participants can access an online metadata submission form via http://data.iutahepscor.org to submit metadata describing datasets that they know about. Users can also view details of datasets that have been submitted by others. The database of metadata that we compile will enable us to both prioritize efforts for providing access to specific datasets through the iUTAH Modeling and Data Federation and will serve as an initial corpus of data that we can use to develop data discovery and access services needed by iUTAH partners.

Sensor Data Management

One of the immediate goals of the iUTAH CI team is to support management of the streaming sensor data from the iUTAH aquatic and terrestrial monitoring sites that will be installed as the GAMUT network is built. The CI Team is assisting in the planning for telemetry connections to each of the iUTAH monitoring sites as well as implementing tools that will facilitate the automated loading of the streaming sensor data into relational databases where they can be more easily managed by the iUTAH watershed technicians and ultimately and shared on the Internet. Much of the required sensor data management functionality is being implemented using existing tools from the CLM4SI Hydrologic Information System.

Extended Data Models

This project is funded through EPS - 1205672. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Model and Data Inventories
Information Entropy

Example of the normal degradation in information content associated with data and metadata over time ("information entropy").

http://dx.doi.org/10.1016/j.ecoinf.2005.08.004.
What if instead?

- Curated data published in a data repository
- Paper using data is published
- Data annotated by additional users
- Data synthesized and leads to another publication

Information Content of Data and Metadata vs. Time
Support for Sharing and Collaboration

- **Collaboration**
  - Datasets as shared “social objects”
  - Formation of collaborations
  - Social networking features

- **Requirements**
  - What data resources are available? (metadata catalog)
  - How do I get them? (search interface)
  - How can I share what I have? (data repository)
Accessing Agency and National Data Networks

• Challenges
  – Multiple sources, websites, systems
  – Multiple data formats, schemas, semantics
  – Building consistent metadata to support data discovery
  – Prioritizing where to allocate our resources

• Solutions
  – Partnerships!!!
Collaborative Data Sharing
cloudshare.iutahepscor.org

http://owncloud.org/
Data Publication in National Networks

• Collaboration with:
  – CUAHSI Data Center
  – HydroShare
  – NSF DataONE Network

• Focus: publishing data products in a way that they can be cited and easily accessed
HydroShare: Collaborative Sharing of Data and Models

- Collaboration environment and social media website
  - Social objects – data, models, digital content
  - HydroShare aims to change the way we do science
  - We are working to make collaboration easier
    - Sharing data
    - Sharing models and other research products
    - Providing new communication and social media capabilities

- 5 Year, $5 Million collaboration among USU, RENCI, BYU, Purdue, U. Texas, San Diego Supercomputer Center, Tufts, U. of North Carolina Chapel Hill, U. of South Carolina

Support: OCI 1148453
**ABSTRACT**
Time series of water quality sensor data in the Little Bear River, Utah, USA.

**KEYWORDS**
Temperature, Dissolved Oxygen, pH, Specific Conductance, Turbidity
<table>
<thead>
<tr>
<th>Type</th>
<th>Title</th>
<th>Date</th>
<th>Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>☐</td>
<td>Little Bear River ODM Database</td>
<td>6/6/2012</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>Little Bear River Sites</td>
<td>6/6/2012</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>Watershed Delineation Workflow</td>
<td>6/6/2012</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>Little Bear River SWAT Model</td>
<td>6/6/2012</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>Little Bear River DEM</td>
<td>6/6/2012</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>Little Bear River Excel File</td>
<td>6/6/2012</td>
<td></td>
</tr>
<tr>
<td>☐</td>
<td>Little Bear River Word .doc</td>
<td>6/6/2012</td>
<td></td>
</tr>
</tbody>
</table>

Results 1 - 10 of 100  << 1 2 3 4 5 6 7 8 9 10 >>
Resource Details

**Resource Type:** ESRI Shapefile

**Created by:** Jeff Horsburgh

**Created:** 6/10/2012

**Keywords:** observations catalog, data cart, little bear river, utah, CUAHSI, Water, Utah State University, Continuous Monitoring, Water Quality, Streamflow.

**Size:** 250 KB

**Extent:** Left: -111.95 Right: -111.79 Top: 41.73

---

**Resource Description**

**Abstract:** Utah State University is conducting continuous monitoring within the Little Bear River watershed of northern Utah, USA to investigate the use of surrogate measures such as turbidity in creating high frequency load estimates for constituents that cannot be measured continuously. This shapefile contains location of USU’s continuous monitoring sites.

**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.

---

**COMMENTS**

Jeff Horsburgh 11.11.12
This shapefile is really helpful if you want to know where my monitoring sites are located.

---

Jeff Horsburgh 11.21.12
Have you thought about re-writing the script so that it can access data using web services instead?
Keyword: Little Bear River

Resource Type: Geospatial dataset


Sort by: Relevance, Title, Owner, Rating, Date

Title: Little Bear River Sites
Shared by: Jeff Horsburgh

Title: Little Bear River DEM
Shared by: David Tarboton

Title: National Land Cover Dataset
Shared by: United States Geological Survey

Title: National Elevation Dataset
Shared by: United States Geological Survey

Results 1 - 10 of 100  << 1 2 3 >>
Groups You Created | Results

Name: Little Bear River Research Group
Description: This group is studying water quality in the Little Bear River, Utah, USA.
Created by: Jeff Horsburgh

Name: CUAHSI
Description: This group shares resources related to hydrologic science.
Created by: Rick Hooper

Name: Little Bear River Research Group
Description: This group is studying water quality in the Little Bear River, Utah, USA.
Created by: Jeff Horsburgh
Next Steps

• Accessing agency and national datasets
  – Developing tools for automatically accessing data
  – Partnerships with data providers

• Supporting data discovery and access across iUTAH and external data sources
  – Mediating across sources, formats, semantics

• Enhanced tools for collaboration and sharing of models and data

• Data publication within national networks
  – publishing data products in a way that they can be cited and easily accessed
Questions?

Jeff Horsburgh
jeff.horsburgh@usu.edu