

LTER Production Workshop Proposal  
SiteDB restructuring and enhancement to support the  
Integration of ClimDB/HydroDB and StreamChemDB into the NIS

Don Henshaw (AND)

18 November 2011

Summary: ClimDB/HydroDB (ClimHyDB ), a data harvester and central relational database, has been largely successful with over 40 participating LTER and USFS sites, 11 million meteorological and streamflow observations, and 30 user sessions per day on average. Sites periodically harvest daily data as it becomes available, and extended metadata descriptions (e.g., contacts, site locations, site characteristics, watershed treatment histories, station locations, and measurement methods and histories) are populated through web forms. StreamChemDB is under development and includes many similar extended metadata features (e.g., basin characteristics, stream sampling methods, lab methods, etc.) structured in a relational database. SiteDB is another network database populated through web forms that was intended to provide a common view of LTER site descriptions and study locations and general descriptions for research themes in common across sites, e.g., climate.

There is considerable overlap in the content of the ClimHy and StreamChem extended metadata with SiteDB, particularly site locations and descriptions, and there are advantages to a single assembly of this type of extended metadata within SiteDB. Advantages of using SiteDB include a) site descriptions and other supporting metadata would be listed in only one place and be consistent and sharable across value-added data compilations such as ClimHy and StreamChemDB, b) display of all measurement sites with these described features would be readily viewable online in a common format and c) further enhancement of LTERMaps would be possible to enable development of map-based GUI applications for data discovery and access. Web forms will be relied upon to populate SiteDB given that these descriptive data change infrequently. Web services will be planned to package this SiteDB metadata into EML as a means of integrating the ClimHy data harvests into the PASTA architecture of the NIS. It will be advantageous to use EML as a means of documenting individual ClimHy data harvests and a first step towards replacing the current exchange format with PASTA-based solutions for populating ClimHY.

Scope of Work:

The primary emphasis will be two-fold:

- 1) Review the current SiteDB data model and modify or extend the design to accommodate the extended metadata features present in the ClimHy and StreamChem data models. Potentially a separate theme or module would exist in SiteDB for each value-added database.
  - a. Eliminate redundancy in content elements among SiteDB and other value-added data models to assure consistent study site locations and descriptions
  - b. Assure that any design changes will accommodate mapping applications planned in LTERMapS. The goal is to assure that a graphical interface could quickly display and access participating sites and stations and associated data for any value-added data set
  - c. Assure that the modified design will facilitate delivery of this information to a web portal or in populating the CUAHSI Observations Data Model (ODM)
- 2) Make recommendations regarding the placement of extended metadata features within EML. The intent is that a web service could be employed to extract relevant metadata from SiteDB to populate EML to accompany a harvest into ClimDB or HydroDB.

- a. Construct a general EML template for organizing SiteDB metadata for ClimDB/HydroDB data harvests that follows LTER EML best practices.
- b. Outline web service requirements for populating this EML template

The scope of work will also consider a) an EML template for a general ClimDB data harvest exchange format, b) necessary changes to SiteDB web forms to accommodate design changes, c) web service requirements for populating the CUAHSI ODM and general wrapping of SiteDB with web services, and d) which RDBMS would best meet the needs for web development and spatial database compatibility.

#### Workshop Products:

- ❖ Complete redesign of the SiteDB data model and associated web forms
  - Implementation of the new model would likely proceed after this workshop but a plan of action for completion will be established
- ❖ EML template for ClimDB/HydroDB data harvests
  - Example templates will be established at the workshop
- ❖ Web service requirements for wrapping SiteDB to populate the EML ClimHy template
  - A plan of action for building this service will be developed
- ❖ Plan for populating the CUAHSI ODM with ClimHy data
  - The plan will describe the necessary steps needed to participate with the Hydrological Information System (CUAHSI-HIS) and build collaboration with CUAHSI.
- ❖ Recommendations for building map-based tools for discovery and access of value-added data
  - These tools will be compliant with the LTERMapS planned development

Participants: The work will require 5-6 participants not including LNO members in a 2.5 day workshop.

Personnel needs include: (listed names are just suggestions so far)

- LNO members (i.e., James Brunt, Yang Xia)
- 1-2 IMs with strong interest in ClimDB/HydroDB/StreamChemDB development (i.e., Don Henshaw)
- 1 member from the EML best practices committee
- 1 IM with web service development skills
- 1-2 members of the LTERMapS committee (i.e., Jamie Hollingsworth, Aaron Stephenson)
- 1 CUAHSI representative for one day (i.e., Jeff Horsburgh)

#### Background Material:

Jun 2011 VirtualWaterCooler notes: <http://im.lternet.edu/node/887>

Jun 2011 VirtualWaterCooler powerpoint:

[http://im.lternet.edu/sites/im.lternet.edu/files/ClimDB\\_integration\\_path.pdf](http://im.lternet.edu/sites/im.lternet.edu/files/ClimDB_integration_path.pdf)

IMC 2011 Breakout Group notes:

[http://im.lternet.edu/sites/im.lternet.edu/files/Session2\\_ClimDB\\_integration\\_breakout.pdf](http://im.lternet.edu/sites/im.lternet.edu/files/Session2_ClimDB_integration_breakout.pdf)

ClimHy schema: <http://climhy.lternet.edu/schema.html> descriptors:

<http://im.lternet.edu/sites/im.lternet.edu/files/descriptors.xls>

Current SiteDB schema, the original SiteDB schema (Baker et al. 2000), and StreamChemDB schemas can be found as attachments: <http://im.lternet.edu/node/887>

Budget: \$6600. This budget is based on 6 people traveling to this workshop, which is planned for Albuquerque, using the Network-recommended average of \$1100 per participant.

Justification:

Albuquerque is the most logical venue for this workshop with two planned Albuquerque-based LNO members attending. Opportunities to conduct this workshop in conjunction with another meeting will be considered if costs can be reduced. Five LTER participants (not including LNO members) will assure adequate knowledge skills are covered for the proposed tasks. An additional CUAHSI participant for one day will assure collaboration with CUAHSI-HIS development efforts.