LTER Production Workshop Proposal

SiteDB redesign to accommodate ClimDB/HydroDB and StreamChemDB

Don Henshaw (AND)

9 December 2011

Summary: This proposed workshop will review the current design of SiteDB and recommend design changes to facilitate the integration of value-added databases, ClimDB/HydroDB and StreamChemDB, into the Network Information System (NIS). The work will focus on the redesign of SiteDB and the development of an EML template to accommodate descriptive metadata from these value-added databases.

Background: SiteDB is a network database that is intended to provide a common cross-site view and simple access to LTER site descriptions, study locations and common research themes, e.g., climate, and includes links to common site resources such as personnel, data, and publications. 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.

There is considerable overlap in the extended metadata content of ClimHyDB and StreamChemDB with metadata included in SiteDB, particularly site locations and descriptions. There are advantages to a single assembly of this type of extended metadata within SiteDB including the following: 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. While recommendations for web services will not be considered here, it is assumed that web services following guidelines of the IMC web services working group will provide an alternative to web forms as a means to populate SiteDB, and additional web services will be employed to package SiteDB metadata into EML data packages to accommodate ClimHy data harvests. 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 these type of cross-site, value-added databases.

Scope of Work:

The primary emphasis will be two-fold:

1. Review the current SiteDB data model and outline necessary modifications or extensions to 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.
   1. Eliminate redundancy in content elements among SiteDB and other value-added data models to assure consistent study site locations and descriptions
   2. 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
   3. 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 (requirements for this web service will not be developed in this workshop).
   1. Construct a general EML template for organizing SiteDB metadata for ClimDB/HydroDB data harvests that follows LTER EML best practices.

Workshop Products:

* Complete plan for redesign of the SiteDB data model
  + 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
* 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 participants not including LNO members in a 2.5 day workshop. Personnel needs include: (listed names are suggestions)

* LNO members (i.e., James Brunt, Yang Xia)
* 2-3 IMs with strong interest in ClimDB/HydroDB/StreamChemDB development (i.e., Don Henshaw)
* 1-2 member(s) from the EML best practices committee
* 1-2 member(s) of the LTERMapS committee (i.e., Jamie Hollingsworth, Aaron Stephenson)

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>
* IMC 2011 Breakout Group notes: <http://im.lternet.edu/sites/im.lternet.edu/files/Session2_ClimDB_integration_breakout.pdf>
* ClimHy schema: <http://climhy.lternet.edu/schema.html>
* ClimHy metadata descriptors: <http://im.lternet.edu/sites/im.lternet.edu/files/descriptors.xls>
* The 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: $5500. This budget is based on 5 people traveling to this workshop, which is planned for Albuquerque, using the Network-recommended average of $1100 per participant. No funding is allocated for participation of LNO participants.

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.