ClimDB/HydroDB Recommendations - 2020-04-06

To: LTER/USFS community

From: LTER/EDI/USFS working group on next-generation system for meteorology/hydrology data

Background

ClimDB/HydroDB https://climhy.lternet.edu is a data warehouse that provides uniform access to common meteorological and streamflow measurements. Through a single portal data from different sites are accessible and simple graphing capabilities are provided. Participating sites are primarily LTER and USFS stations. Although the ClimDB/HydroDB data warehouse accepts 14 measurement parameters in a daily time step, precipitation, air temperature and stream discharge account for more than 75% of the data. The primary use of the data has been for research, synthesis, and education (e.g., K-12, undergraduates).

The ClimDB/HydroDB was migrated to the LTER Network Office at UNM in 2010, but further upgrades to this custom application are not planned and the infrastructure is antiquated. Consequently, the working group reevaluated the needs for harmonized multi-site and multi-agency climate and hydrologic data, the infrastructure required to collect and store these data, and considered new technologies and workflows based on current data storage models. Feedback from LTER and USFS scientists regarding requirements for a system holding meteorology and hydrology data indicated the following valuable considerations:
   1. Harmonized meteorological and hydrological data
   2. Time resolutions as high as 15 minutes are desirable
   3. Mechanisms to query and retrieve data from several sites
   4. Ability to easily merge data streams across sites, as well as merge with other data types (e.g., soil types, biogeochemical parameters)
   5. Dashboards, basic plotting, graphing, and visualization functions
   6. ClimDB/HydroDB is regularly used by some sites

Recommendation

The working group evaluated several existing options and concluded that *ClimDB/HydroDB could be replaced by the CUAHSI (Consortium of Universities for the Advancement of Hydrologic Science, Inc., https://cuahsi.org) data warehouse*. CUAHSI represents more than 130 U.S. universities and international water science-related organizations, and receives support from the NSF to develop infrastructure and services for data access, analysis, and collaboration. It already delivers similar data from other sources, such as USGS streamflow gages. Additionally, CUAHSI's data model could be used to *standardize and harmonize LTER's meteorological and hydrological*

data. The CUAHSI HydroClient https://data.cuahsi.org/ would provide the search interface and basic graphing capabilities desired by data users. We view LTER contributions to CUAHSI as a big opportunity for sites to make their climate and hydrology data more accessible to a broader community.

We advocate a two-part migration approach, which are independent and could be conducted simultaneously:

Part A. Archive ClimDB/HydroDB content in the EDI data repository (anticipated Summer 2020), after which ClimDB/HydroDB will be sunsetted.

Part B. LTER sites develop new workflows to convert raw data into the CUAHSI data format. Raw data are archived in the EDI data repository. CUAHSI formatted data are archived in EDI as well as CUAHSI.[1]

Process

Part A: Archive climDB:

1. All data from the current ClimDB/HydroDB database will be exported into flat files on **2020-07-01**. Decisions about these structures are still being examined (by station, parameter, and aggregation - daily, monthly, annual). Create EML metadata and upload the package into PASTA. As much metadata as possible will be pulled from the current ClimDB/HydroDB database.
2. Before the final archive of ClimDB/HydroDB on **2020-07-01**, contributors may be requested to augment metadata and optionally, update data. (https://climhy.lternet.edu/siteDB.pl?use=Update).
3. The archived data from ClimDB/HydroDB packages in PASTA will be converted to CUAHSI ODM data format and uploaded to CUAHSI as an independent data source.

Part B: Ongoing contributions:

1. LTER sites develop workflows to convert raw data into CUAHSI ODM, similar to the past procedure for adding data to ClimDB/HydroDB. EDI will mediate the process by:
   a. Providing documentation https://github.com/lter/Clim-HydroDB-2.0).
   b. Assisting with conversions to CUAHSI for raw data already in an EML data package in EDI with R scripts.
2. Sites submit converted data to EDI, in addition to raw data.
3. Sites submit converted data to CUAHSI.
4. The preferred time resolution is 15 minutes; however, other resolutions are acceptable.
5. Detailed documentation of the ClimDB/HydroDB working group decisions can be found at: https://github.com/lter/Clim-HydroDB-2.0/blob/master/meeting_notes/2019-03_workshop_ABQ/workshop%20decisions%20documentation.txt

---

[1] Advantages of data being held in EDI in addition to CUAHSI: Data are citable in (doi); easily cataloged alongside other LTER data packages; ability to build and apply tools to data packages.