LTER IMC VTC

2019-04-08

notetaker: Stevan

Suzanne opened the call for comments or questions about the IMC/ESIP meeting.

Gastil and Tim resolved that Gastil will submit the proposal for the metabases topic.

Margaret inquiring as to how many ESIP sessions IMC members are planning to submit. We need to resolve this. Suzanne indicated that IM Exec will send an email seeking input and inquiring as to what sessions IMs are planning to submit.

Main topic for today: Suzanne, Don, Wade, and Margaret presenting results of the ClimDB next generation meeting

Suzanne gave an overview of the workshop objectives. Workshop attended by EDI, LTER Information Managers, and scientists.

Don provided a summary fo the workshop: - Current contents of ClimDB will be archived with EDI. - Discussed at the meeting the best form for our climate and hydrologic data. CUASHI observation data model was identified as a suitable format. - Output from the workshop will be sent to the ClimDB committee for comment. At least one additional meeting is planned.

Don gave a brief history of ClimDB (Don): - A common format for hosting standardized climate and hydrological data. - A harvester was developed to harvest and format data, could also access data from USGS and other resources beyond just LTER sites. - ClimDB featured tools to facilitate data acquisition, data mining, visualization, etc. - There are 45 participating sites (LTER, USFS, USGS) - 389 total measurement stations: 189 meteorological, 200 stream gauging - 21 daily measurement parameters - $>$ 10 million daily values - a great platform but we need to move forward with current technologies

Plan for existing data (Don): - Archive existing data with EDI. - Sites can update their data before it is archived with EDI

Plan for future data (Don): - Prepare existing data in the CUASHI format - ClimDB parameters have already been mapped to CUASHI - The data formatted for CUASHI would also be archived with EDI

Don clarifying from a question/comment by Lindsey that data providers will have an opportunity to update their contributions to the existing ClimDB. The deadline for this is to be determined.

Don noted that Lindsey has been active in trying to promote the data coming from the experimental forest.

Wade providing more detail about CUASHI

CUASHI was one of several potential targets for these data. It is open to contributions, and the data model works well for the existing data.

CUASHI is a long-running, stable, NSF-funded project. It is a cyberinfrastructure for sharing point-based data over time. Time series is a combination of a site, methods, source, and a quality level. Built on a relational model. Currently on version 1.1. A future version may add GIS features. It is a skinny data model. It is based on a series of metadata tables that describe the content for the observations in the skinny data model, geographic coverages, methodology, details about variables, QC levels, and extensive controlled vocabularies, among others - a very thorough and comprehensive system.

Requires variables, methods, sites, sources, quality control, and data values for contribution. As Don mentioned, these are already mapped between ClimDB and CUASHI. Wade went details about each of the required elements - Don and others have already done a lot of the potential harvesting from ClimDB. The real work is in the data values table which links the data with all the provided (and required) metadata. CUASHI has a metadata template in the form of a spreadsheet for providing metadata.

CUASHI's spreadsheet template would not scale well for ClimDB. However, there are features and aspects of the GCE toolbox that would be a good place to start for managing the transition to CUASHI. Much of the work will be putting together metadata details of individual sites. GCE toolbox could also be helpful to translating the existing ClimDB to EDI. Sites could use the tools to address ClimDB from their own sites for contributing to EDI.

Margaret provided some high-level summary slides regarding the EDI side of this

- Level 0: raw
- Level 1: based on predefined model
- Level 2: derived data

EDI would used a harmonized approach that would treat ClimDB much like ecocomDP. The advantage here is that because they are standardized, data could easily be harvested from the L1 state. In this case, L1 would be the predefined ODM model. Level 2 would be adding the data to CUASHI.

Still deliberating as to whether existing ClimDB would be archived with EDI as is or in the ODM format for CUASHI input. GCE Toolbox will be used for this. But how do we organize the data, e.g., one for each contributor, what is the refresh schedule, and there may be additional information requested.

Phase 2 will consider ongoing ODM-style datasets (L1). Future workflows are likely to vary by source (e.g., USGS data will be different from an LTER site data).

Suzanne/Margaret opened the floor to comments/questions:

Lindsey adding that she was tasked with spending time with the nuances of CUASHI. She noted that the tools are very clunky, limitations on uploads/downloads, visualizations, etc. So, we are interested to follow-up with CUASHI about their next steps.

Stevan asking what he needs to do? At this point, people not directly involved should be thinking about the metadata that should be requested and archived.

Tim inquired about how to move the conversation forward. Best approach now is to email someone on the working group. Margaret will set up a Slack channel for this topic.

Wade suggesting that there will be a short transition from theoretical to a working model.

Corinna responding further to Stevan: think about the sources of data, from how many weather stations would a site like to contribute data. Corinna noted that many investigators feel having aggregated weather data for the LTER is very important.

Lindsey mentioned that we need to consider branding. But Wade adding that organizational scope for finding data is very important.

Suzanne inquiring if some steps, like providing the L0 data could be skipped. Margaret and Corinna noting that L1 format is most critical, but such issues may be site-specific and/or case-by-case.