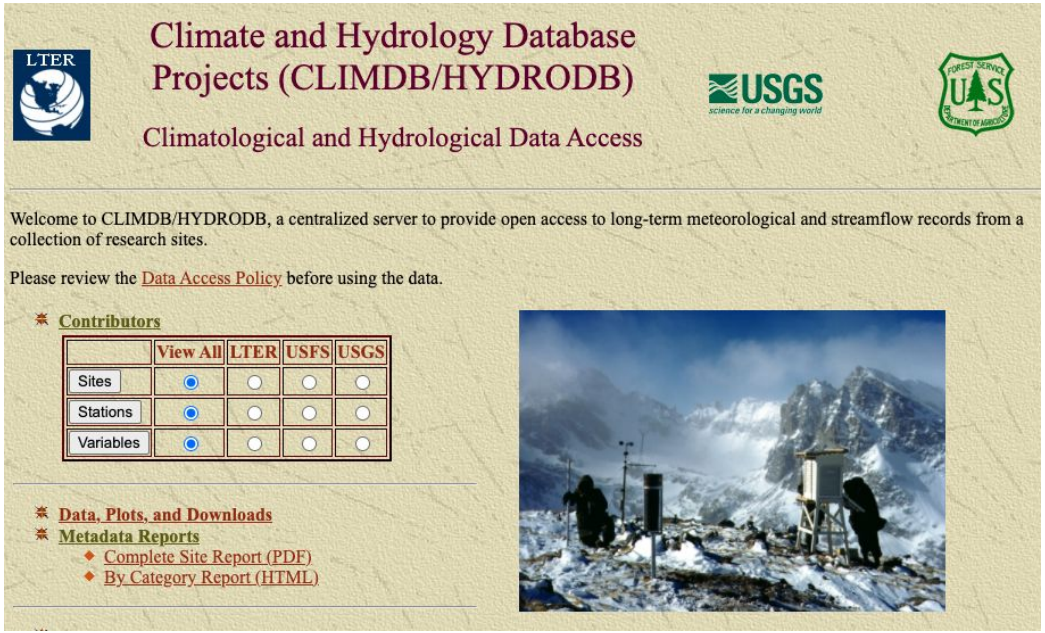
Update: Harmonizing meteorological and hydrological data in the EDI data repository

15 December 2020

Background

https://climhy.lternet.edu



https://github.com/lter/Clim-HydroDB-2.0/blob/master/planning_documents/Plan_climDBv2.pdf

Background

Replacement system needs …

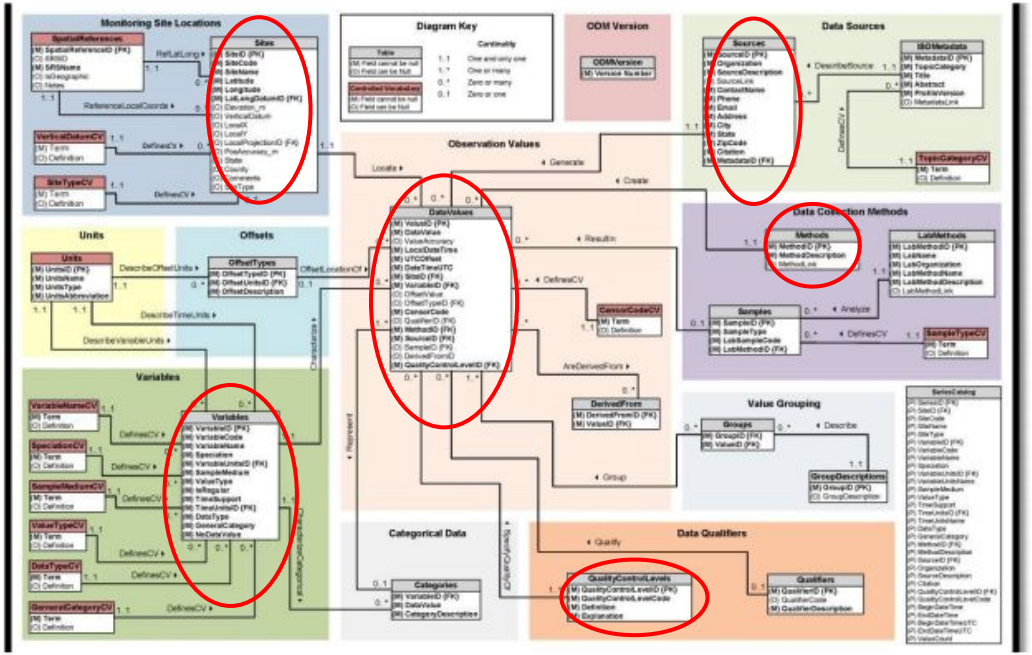
1. Data in one format (harmonized, analysis ready)
2. Higher frequency
3. Mechanisms to query and retrieve data from several sites
4. Ability to easily merge data across sites, or other data types
5. Dashboards, basic plotting, graphing, and visualization functions

https://github.com/lter/Clim-HydroDB-2.0/blob/master/planning_documents/Plan_climDBv2.pdf

CUAHSI
Universities Allied for Water Research

CUAHSI
Hydrologic Information
System (HIS)
ODM 1.1

https://data.cuahsi.org/



https://github.com/lter/Clim-HydroDB-2.0/blob/master/planning_documents/Plan_climDBv2.pdf
http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf

EDI Objectives

1. Archive ClimDB/HydroDB database in EDI
   a. 23 tables, DB native format
   b. ERD: https://climhy.lternet.edu/schema.html
2. Convert ClimDB data to CUAHSI ODM format, in EDI
   a. 59 data packages (1/contributor)
   b. 6 tables per package

A

3. Provide
   a. guidance for CUAHSI contributions
   b. Platform for workflows

B

Progress - Part A

Archiving ClimDB/HydroDB tables

- Available for review, Jan 2021 (staging server)
- Data entities
  - 23 tables
  - ERD image
  - Database zip with DDL

CUAHSI ODM conversion for each site

- Example (Andrews Forest): URL - EDI staging portal
- Data entities
  - 6 “Standard” ODM tables (DataValues, Sites, Sources, Variables, QualityControlLevel, Methods)

Logistics

GitHub for guidance on how to prepare data in CUAHSI ODM, workflows

- https://github.com/lter/Clim-HydroDB-2.0/tree/master/climdb-archive-partA
  - R-code
  - SQL code
  - EML in progress
  - Some data entities

Example Data Package

Weather and hydrological data for H.J. Andrews Experimental Forest, previously held in LTER ClimDB/HydroDB(1949 to 2020)

https://portal-s.edirepository.org/nis/mapbrowse?scope=edi&identifier=680&revision=newest

Data Package Details

Climdb/HydroDB is not a perfect match to ODM. You will want to review the 6 ODM tables and **review and complete the information**, specifically:

Sites.csv: geographic coordinates & datum, elevation
Sources.csv: source description (person, organization)
Methods.csv: methods used for specific data values
QualityControlLevels.csv: types of data (raw, processed) or other
Variables.csv: “NoData” codes
DataValues.csv: UTCOffset, linkages to all other tables

CUAHSI Contributions

LTER sites' tasks

- **Finalize ODM tables** with information not available from ClimDB/HydroDB.
- Edit the **EML** *
- TALK TO US (Susanne and Margaret)

EDI's tasks

- **Publish** up to 59 packages from ClimDB
- **Advice and support** in completing your ODM tables
- **Hackathon** * to review and fill ODM tables

***Question:** Should we set a deadline for finalized packages? This would make EDI's tasks easier.*

CUAHSI Uploading Steps

1. Create a publishing account
2. Verify your account using the activation email
3. Format your data
4. Upload your data
5. Request Publication

https://github.com/lter/Clim-HydroDB-2.0/tree/master/CUAHSI_documentation

Questions about your CUAHSI account and uploading
CUAHSI support: help@cuahsi.org

Questions about the ODM standard format and using workflows
EDI support: support@edirepository.org

EDI Objectives

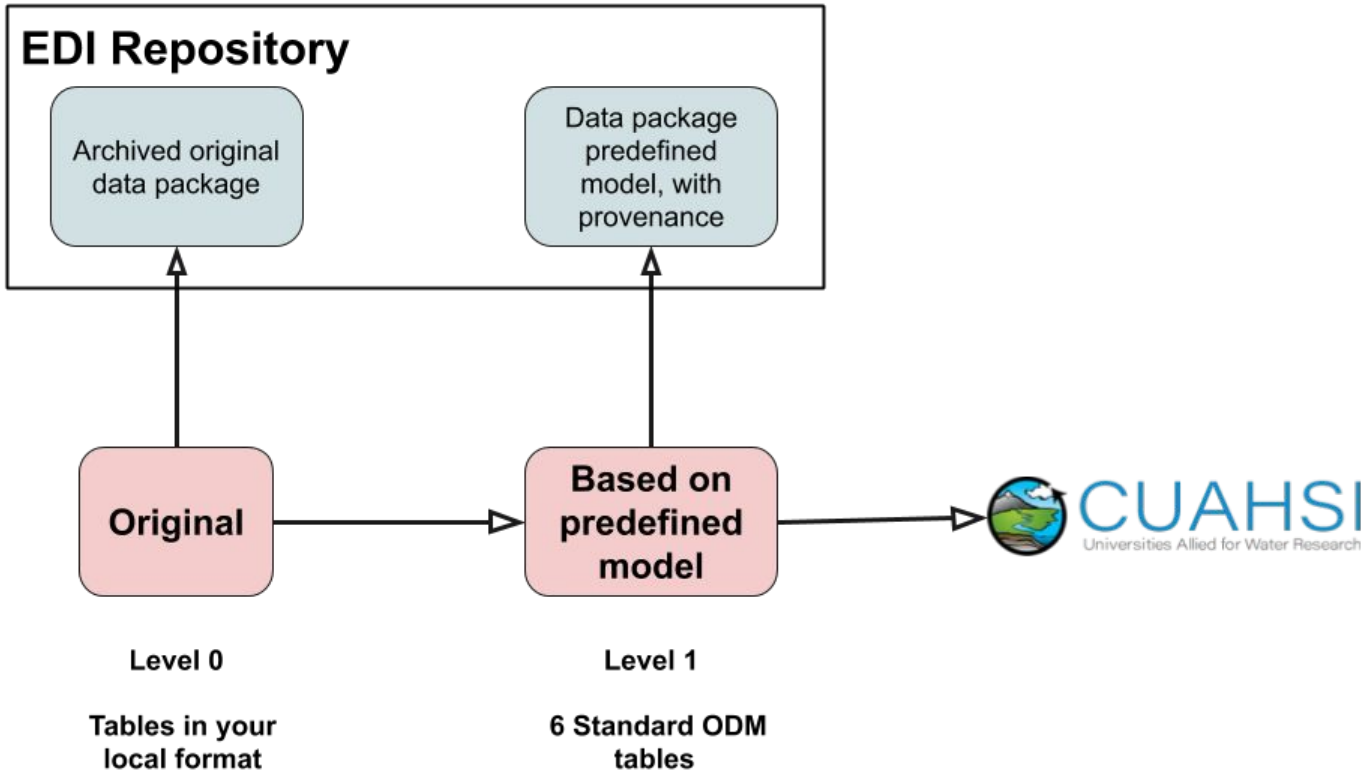
1. Archive ClimDB/HydroDB database in EDI
   a. 23 tables, DB native format
   b. ERD: https://climhy.lternet.edu/schema.html
2. Convert ClimDB data to CUAHSI ODM format, in EDI
   a. 59 data packages (1/contributor)
   b. 6 tables per package

A

3. Provide
   a. guidance for CUAHSI contributions
   b. Platform for workflows

B

EDI Workflows



CUAHSI Contributions

LTER sites' role

- **Create your own CUAHSI account**
- Plan your CUAHSI contributions, e.g.,
  - Package from ClimDB/HydroDB (edited if necessary)
  - Your L0 (which may differ from what was in ClimDB/HydroDB)

EDI's role

- Advice on ODM format, CUAHSI practices
- Support for your L0 -> L1 workflow using PASTA event notifications
  - Hackathon

Comments and Discussion

Hackathon ideas

Harvester for CUAHSI, using their API
Issues: data ownership,

Script to keep ODM-style datasets up to date

Scripts to use ODM-styled datasets directly from EDI

For ongoing data, L0 EML to L1 EML (similar to ecocomDP conversions)