ClimDB/HydroDB (ClimHy) Integration

- ClimHy has been migrated from AND to LNO and will remain status quo in 2011
  - Public page (http://climhy.lternet.edu/)
  - Participant page (http://climhy.lternet.edu/harvest.html)
  - Databits Fall 2009 (http://databits.lternet.edu/fall-2009/climdbhydrodb-climhy-database-migration-lno)
  - Database schema (http://climhy.lternet.edu/schema.html)
- Integration Path document (Brunt, Servilla)
  - http://intranet2.lternet.edu/content/climdbhydrodbecotrends-network-database-integration-path
- ClimDB integration committee
  - Current commitments: Don, James B, Yang Xia, Suzanne
  - Looking for 2-3 additional members

ClimDB/HydroDB (ClimHy) Status

- Status of populated data
  - Primarily populated with air temperature, precipitation, and streamflow (daily values)
- Status of current participation (2 June 2011)
  - 26 LTER sites participating
    - 16 sites with climate data current (within 1 year)
    - 5 sites with climate data current (within 1-2 years)
    - 3 sites with climate data 2-4 years out of date
    - 2 sites with climate data >4 years out of date
  - 15 Non-LTER USFS sites participating
    - 3 sites with streamflow data current (within 3 years)
    - 12 sites with streamflow data >3 years out of date
  - 3 Taiwan sites participating
    - 2 sites with climate data current (within 2 years)
    - 1 site with climate data 2-4 years out of date
- Status of metadata (e.g., ClimDB specific metadata)
  - 9 sites with demonstrated effort to populate significant metadata at station and parameter level
  - 10 sites with demonstrated effort to populate some metadata
  - 7 sites showing virtually no metadata

Description of populating routines

- Sites manually or programmatically prepare harvest files of daily data from local data sets
- All sites have the ability to go to the participant page to "push" an update. The data can be harvested from one of two user-specified URLs
- 4 sites are on a weekly schedule in which ClimHy "pulls" climate data from the specified site URL
- 15 sites are on a weekly schedule in which the GCE Toolbox queries ClimHy, pulls USGS data from web sites, resamples and reformats, and pushes data into ClimHy
  - NOAA NCDC station data can be harvested
  - http://gce-lter.marsci.uga.edu/public/im/tools/usgs_harvester.htm
- The Andrews incorporates a ClimHy web service into its data preparation – the service recalls the date of the last harvested value by station and measurement parameter
- Metadata can only be entered manually through web forms

ClimHy data summary statistics

3 June 2011 (Yang, Remillard, Henshaw)

- 30 ClimHy web sessions per day on average
- Data use
  - Total Downloads since inception of web site (2003):
  - 20,490 downloads, plots, or views
    - 8741 data sets downloaded
    - 9453 plots created
    - 2295 views of data
- Downloads by year:

Year	downloads	plots	views
2003	309	1240	291
2004	267	566	98
2005	717	829	191
2006	1886	978	335
2007	946	816	210
2008	1259	888	347
2009	1150	946	281
2010	1969	3042	525

- Participation
  - Over 11 million daily values in our database.
  - 364 measurement stations
  - 26 LTER sites
  - 3 ILTER (Taiwan) sites
  - 22 USFS sites
  - 15 sites with USGS stations
  - 21 total measurement parameters

Possible Integration Pathways
2011-2012

- Create EML for current harvest data structure
  - Currently a similar format for each site (CSV file)

    !Lter_site, station, date, field1, flag_field1, field2, flag_field2,field3, flag_field3, field4, flag_field4
    ABC,MY_STATION,19970228,111.1,,222.22,E,333.3,,444.4,
    ABC,MY_STATION,19970304,,,,,,,34,Q
  - Order of field names and number of header lines can vary
  - USFS and other participating sites will likely need a standard EML template
- Harvest original (level 1) data and use workflows in PASTA to create ClimDB compatible data
  (e.g., convert 15 minute data to daily)
- Harvest processed (level 2) daily data into PASTA

Possible Metadata Integration
2011-2012

- Merge specific ClimDB metadata into existing EML elements
  - See uploaded excel file descriptors.xls to see metadata framework
- Prepare this metadata as if it were a data file, describe with EML, and harvest separately from the ClimHy data
  - For each station there could be one EML file, one table of observations, and one table of site of site documentation
- Divert some of the metadata into SiteDB and build and populate modules for site descriptions, gaging and meteorological station descriptions, etc.
  - StreamChemDB will require much of the same information

Proposed Timeline, 2011-2015

- 2011
  - Initiate a new ClimDB integration committee
    - Plan breakout time at annual meeting
    - Team with scientific workflow production workshop (late 2011)
  - Consider approach… develop EML document requirements
- 2012
  - Sites will develop EML for ClimHy prepared data based on these requirements
    - PASTA will trigger data harvests into the NIS
  - LNO will develop a workflow script to load data from PASTA into the ClimHy database
  - Sites begin to develop workflows to load common site climate data into ClimHy
- 2013
  - LNO with help from IMC will refactor the existing ClimHy ingestion engine into RESTful web services
- 2014-2015
  - Tools will be prototyped in the NIS data portal to produce data products as prioritized by network committees

Current StreamChemDB database design