16 Oct 2001

# ClimDB Progress Report

**Introduction:**  A new version of ClimDB is now ready and features a greatly enhanced harvest feature as well as prototype metadata entry forms. This newest implementation of ClimDB allows participating sites to trigger a harvest of their own site’s data from the central site webpage. ClimDB now sports two web pages: a new participant page as well as the original public web page. Both pages currently reside at the Andrews LTER (Oregon State).

**Participant page:**

[**http://www.fsl.orst.edu/climdb/harvest/harvest.htm**](http://www.fsl.orst.edu/climdb/harvest/harvest.htm)

**Public page:**

[**http://www.fsl.orst.edu/climdb**](http://www.fsl.orst.edu/climdb)

Participant page features:

* New user guide describing exchange formats, variables, metadata, and detailed harvest instructions and error messages
* Dynamic participating site list showing all LTER sites, climate PI’s and other researchers, data set and metadata contact persons
* Harvest trigger allows do-it-yourself harvesting of data
* Metadata description and metadata entry forms
* Personnel form to enter data set contacts and researchers

The self-harvest trigger allows the success or failure of the harvest and subsequent ingestion into the database to be known immediately. An error log will be posted to the screen as well as automatically mailed to the data set contact person(s) for the site and the ClimDB administrator. This self-harvest mechanism will allow necessary changes or corrections be made directly by the site in an iterative harvesting process until successful. Sites will be solely responsible for the harvest of their own data, and central harvesting of all sites will end.

We have established a mailing list of the lead investigators, other researchers, and data set contact persons (climdb\_all@fsl.orst.edu), and a second mailing list of data set contact persons for distributing more technical information about the harvest process (climdb@fsl.orst.edu).

**Required Action:** We need each site to please do the following:

1. Review listed personnel for your site (Click on “Sites in the database” from participant page): (<http://spiraea.fsl.orst.edu/climhy/show_sites1.asp>)
2. Let us know of any changes in personnel. A web form is available to enter new personnel information at (Click on “Adding personnel” from participant page - Internet Explorer works better than Netscape) <http://www.fsl.orst.edu/climdb/harvest/frmPersonnel.htm>, or email information to Henshaw@fsl.orst.edu.
3. Let us know your data URL that is the web address of the exchange format file for harvest. A mechanism to allow editing of this URL is planned but not yet implemented.
4. Let us know the code and name of all meteorological stations you are planning to include in harvest files. This code must be in our database before harvest (codes<=10 characters).
5. Please review the user guide (Click on “User guide” from participant page): <http://www.fsl.orst.edu/climdb/documents/climhydro_userguide.html>
6. Harvest your site’s data: Click on “Harvest Data” from participant page.
7. Enter metadata. Click on “Updating metadata” from participant page.

**ClimDB Metadata Implementation:** Metadata will be collected for each weather (meteorological) station and all individual measurement parameters at each station. Minimal metadata at the Research Area/Site level is also requested. Web forms are established for individual sites to provide this information. Metadata elements have been grouped into descriptor categories for ease of entry and implementation on a web form (Click on “Variable names and descriptions”).

##### <http://www.fsl.orst.edu/climdb/harvest/climdb_descriptors.htm>

1. Research Area Information
2. Meteorological Stations
3. Climatic Measurement Parameters

**Note:** Meteorological station names must be placed into the master database before any metadata entry or data harvest can proceed.

**Comments and Caveats:** While the existing ClimDB database was moved in tact from the Network Office, every site should perform a harvest of their data (click on “Harvest Data” from participant page). This will 1) familiarize you with the new process, 2) ensure your data is current and accurate, and 3) inform you of current errors within your exchange data. Note that several sites currently show errors and warnings on harvest attempts. Please refer to the User Guide for more information and error and warning message interpretation. [Note: you will receive a warning if no header line is provided. While a default header line is provided, this practice is discouraged]

Additionally, it is no longer necessary to continually re-harvest the entire data set each time. Any data that is re-harvested will overwrite the data in the database, but only for that date and that parameter that is re-harvested. That is, it is only necessary to re-harvest data points that have changed. However, it is not a problem to re-harvest the entire data set. Any new values encountered will be appended to the data set. [Note: this new update method prevents sites from “dropping out” of ClimDB on harvest failure. In general, fatal errors on harvest will have no effect on the current database. Also, if it becomes necessary to delete a range of dates that cannot be replaced with corrected data, contact the ClimDB administrator.]

Global QA max-min range checking is performed on all harvested data. Sites can dictate site-specific range values for each station’s measurement parameters used for this check. Once the site enters the max-min values in the metadata for a measurement parameter, these max-min values will be used in lieu of the global values for this QA check.

**Disclaimer:** 1) The metadata entry forms are largely untested. I would welcome getting feedback on anyone’s successes or failures. 2) Netscape is problematic on the personnel entry form, but seems to work fine with most other features including data harvesting and metadata entry. Internet Explorer works fine. 3) The harvest process can take a few minutes. It is probably best to keep the browser running until you see this link to the public data page, “Would you like to go to the [ClimDB data page](http://www.fsl.orst.edu/climdb/)?” However, fatal errors will prevent you from seeing this message and link.

**Comments on future integration and development:** The harvester and supporting database design support both ClimDB and a hydrological database (HydroDB). The ClimDB/HydroDB database and webpages are currently located at the Andrews LTER. Once the harvester, database, and metadata entry forms are tested and considered robust, the database and webpages can be ported back to the NET office. The supporting database structure has been designed for future integration of these research modules with the Network Information System (NIS) and SiteDB. Funding for this integration and long-term maintenance of the database will be required before the ClimDB/HydroDB modules can be ported to the NET Office.

Other enhancements are also necessary. 1) Current development has focused on the harvest engine and not the public data download and graphics page. Both the download interface and graphics interface are generally functional, but do not adequately support the additional new data variables. Recommend both interfaces be rewritten. 2) There is no request tracking form, web counter, or data access policy. 3) There is no mechanism to download metadata. Metadata entry forms have been developed to capture metadata in the database, but no mechanism is provided to produce a downloadable metadata report. 4) Password protection is needed on the self-harvest page. 5) Front-end tools are necessary for maintenance of sites, stations, and personnel.

**Developers:** The current ClimDB/HydroDB Development Team is located at the Forestry Science Lab in Corvallis, Oregon.

Don Henshaw: Clim/HydroDB database administrator (Henshaw@fsl.orst.edu)

Suzanne Remillard: Clim/HydroDB user support (suzanne.remillard@orst.edu)

Kyle Kotwica: Clim/HydroDB programmer/technical support

The U.S. Forest Service Fire Evaluation Monitoring Program is funding this current development of both ClimDB and HydroDB. This effort builds upon development efforts funded by the National Science Foundation and LTER Information Managers in the creation of ClimDB, in particular, Barbara Benson and Robin Stubbs of the North Temperate Lakes LTER site, Richard Dahringer of the LTER Network Office, and Kristen Overturf at the H.J. Andrews site.

Similar web pages have been developed for HydroDB:

<http://www.fsl.orst.edu/hydrodb/harvest/harvest.htm>

<http://www.fsl.orst.edu/hydrodb>