Ingest Workflow and Quality Control for International Tree Ring Data Bank (ITRDB)

Contents

Changelog

2024-01-22: Edit for clarity and consistency; convert document to pdf.
2023-01-10: Update URLs; add information about COFECHA use for quality assurance and reusability.
2019-04-22: Initial creation of this document.

I. Purpose

The International Tree Ring Data Bank (ITRDB) is the world's largest public archive of tree ring data. of fire history data derived from natural proxies. The ITRDB includes raw ring width, wood density, and site growth index chronologies from more than 5,000 sites on six continents. The purpose of this file is to document the data ingest workflow and quality control of ITRDB records.

II. Data Contribution Files

Complete guidelines for contributing ITRDB records are located at:
https://www.ncei.noaa.gov/products/paleoclimatology/contributing-data.

ITRDB data contributions are typically received in the Tucson decadal format rather than in the NOAA WDS-Paleo template. However, the NOAA WDS-Paleo template is used to capture the metadata. The Tucson decadal format is described at:
https://www.ncei.noaa.gov/pub/data/paleo/treering/treeinfo.pdf

III. Quality Control, Curation, and Ingest

When a new tree-ring data collection is received, typically by email, but also possibly on media or via drop box download, the following procedures are followed:

A. Visually inspect Tucson decadal format and test the formatting of the rwl file by making sure the file can be read using the read.rwl() function from the dplR R package (available at: https://cran.r-project.org/web/packages/dplR/index.html).

B. Verify that metadata components are in proper column locations - edit or add headers as needed. Metadata 3-line headers may be included in the file upon submission, but we request that the metadata be contributed separately in the metadata portion of the NOAA/WDS-Paleo template. The template allows the contributor to provide a more complete metadata record than is possible from the 3-line header. The standard metadata header is described at: https://www.ncei.noaa.gov/pub/data/paleo/treering/treeinfo.pdf

C. Run COFECHA statistical cross-dating software for each set of raw tree ring data (Available from U of Arizona LTRR Program Library at: https://www.ltrr.arizona.edu/pub/dpl/). Verify that each collection meets the minimum ITRDB COFECHA thresholds of less than 40% "problem segments" and greater than 0.35 mean series intercorrelation. These standards were proposed in Jeff Lucas' 2005 ITRDB white paper, and adopted by the ITRDB Advisory Committee. If all data meet the criteria, acknowledge receipt to contributor, proceed to ingest data into the ITRDB. If any collection fails to meet the criteria, advise contributor that the data have failed COFECHA standards, and suggest re-evaluation of the cross dating of the samples, with possible re-submission of data following modification.

D. If all data pass COFECHA standards, assign SITE CODES. This identifier code consists of a location code and 3-digit numeric identifier. Location codes are by country except for data from the USA, where codes are by state and are the postal code for the appropriate US state. Examples include AK143 for the 143rd collection archived from the state of Alaska, or CANA112 for the 112th collection from Canada. Codes are stored in the "Code" field in the PC_studies table, and also form the root of the archived data file names. Data managers can use either source to determine the next available code to assign to new data. If multiple species are collected at a location, each species/site combination is considered a single collection, with a separate ITRDB code.

E. Each ITRDB code and all associated data files are entered into NOAA/WDS Paleo Oracle database via the PINGMAN data entry application, or via batch ingest. Each study will consist of one tree ring "collection", a group of samples from a single tree species at a location (i.e., each ITRDB code will correspond to one Study ID record). A minimum of 2 files will be included, one file of raw measurements, file name pattern CODE123.rwl, plus an information file containing COFECHA output and notes, file name pattern CODE123.txt. For quality assurance and to support reuse, the document containing the output of COFECHA is provided with each published dataset. Other measurement types and optional chronologies may increase the file count for the study. Most collections in the ITRDB are total ring width, with a minority of contributions also having additional parameters contributed (Earlywood width, maximum density, etc). Site chronologies are single time series, generated from the raw measurements, and stored in separate files with file naming pattern CODE123.crn. These are optional, and may or may not be contributed along with the raw data they are based on.

F. Data and information files are uploaded to the public FTP server, specifically:
    a. COFECHA output is archived in a text file, with notes and references for the collection, in the online folder: https://www.ncei.noaa.gov/pub/data/paleo/treering/measurements/correlation-stats
    b. Raw measurements RWL files are stored in the online folder: https://www.ncei.noaa.gov/pub/data/paleo/treering/measurements/ (this is divided into subfolders by geographic region)
    c. Chronology CRN files are stored in the online folder: https://www.ncei.noaa.gov/pub/data/paleo/treering/chronologies/ (this is divided into subfolders by geographic region).

IV. Identifiers, Access, and Archive

Each ITRDB dataset ingested into the NOAA WDS-Paleo database receives a unique internal study identifier for tracking purposes, a DOI for permanent data location, and three metadata records in ISO, DIF, and JSON formats. The ISO-19139 record is quality controlled using NCEI's rubric for automated metadata checking.

This ingest process also makes the data accessible to the end user through the WDS-Paleo Dataset Search: https://www.ncei.noaa.gov/access/paleo-search/?dataTypeId=18, as well as other Data Access tools located at: https://www.ncei.noaa.gov/products/paleoclimatology/tree-ring

On a monthly basis, all WDS-Paleo data are placed in the NCEI long-term archive.