Ingest Workflow and Quality Control for International Multiproxy Paleo-fire Database (IMPD) Charcoal-Based Data

Contents

Changelog

2024-01-22: Edit for clarity and consistency; convert document to pdf.
2023-01-10: Initial creation of this document.

I. Purpose

The International Multiproxy Paleo-fire Database (IMPD) is an archive of fire history data derived from natural proxies. The IMPD includes data from charcoal in lake sediment records, and tree-ring based data including tree stand establishment data and fire scars in the annual growth rings of trees. The purpose of this file is to document the data ingest workflow and quality control of charcoal sediment records. The tree-ring based data ingest workflow and quality control is documented at: https://www.ncei.noaa.gov/pub/data/paleo/data_management/impd-tree-based-workflow.pdf.

II. Charcoal-based data contribution files and guidelines

The ingest workflow begins with the contributor sending data to the World Data Service for Paleoclimatology/IMPD. Complete guidelines for contributing charcoal-based IMPD datasets are located at: https://www.ncei.noaa.gov/products/paleoclimatology/contributing-data. New data acquisitions arrive typically by email (paleo@noaa.gov), but also possibly on media or via drop box download.

Charcoal-based data contributions consist of the following data files:

1. **Required:** Fire History Charcoal-Based Data Contribution Template which is located at: https://www.ncei.noaa.gov/pub/data/paleo/templates/impd-charcoal-template.xlsm. The "README-INSTRUCTIONS" tab contains complete instructions on what to submit and entering the metadata and data.

2. **Optional:** Files containing supplemental information.

III. Quality Control, Curation, and Ingest

Data contributions are inspected by the data manager at the start of the ingest process. The data manager will work iteratively with the investigator to work through any issues in the formatting or description of the data set. The data manager and the original data contributor will also do a final review of the data set upon creation of a public-facing landing page for each IMPD data site ingested.

As a first step for quality assurance, the contents of Data Contribution Template are assessed via visual inspection, which includes information provided in the publication manuscript associated with the dataset.

A Paleo Data Manager then creates files to be archived and ingests metadata into the Paleo Oracle Metadata Database as follows:

A. Each new IMPD data site is assigned a unique IMPD Code by the IMPD Data Manager. The first 2 characters of the IMPD Code represent the two letter International Organization for Standardization (ISO) country code (https://en.wikipedia.org/wiki/List_of_ISO_3166_country_codes). The next three characters represent the site code, and the last three characters are a sequential numeric index used to uniquely identify files with duplicate country and site codes.

B. Using Excel macros, the data in each of the three data tabs of Data Contribution Template (charcoal sediments, magnetic susceptibility, and loss on ignition) along with the metadata of the Data Contribution Template's metadata tab are transformed into a NOAA/WDS-Paleo Template file (https://www.ncei.noaa.gov/pub/data/paleo/templates/noaa-wds-paleo-template-instructions.txt).

C. The NOAA/WDS-Paleo Template files to be archived are named according to the assigned IMPD code with the extension based on the variable measured (.txt for charcoal sediment, .mag for magnetic susceptiblity, and .loi for loss on ignition).

D. Each of the NOAA/WDS-Paleo Templates created are then run through the WDS-Paleo QC checker to ensure that metadata are filled out correctly and contain valid information, and that the data table is formatted correctly.

E. Using the metadata and data of the Data Contribution Template as input to the Paleo Oracle Metadata database and associated PINGMAN ingest application, all database metadata records are created. These records include investigators, funding, and publications, site and location information.

F. Data files are placed on the public FTP server.

IV. Identifiers, Access, and Archive

Each IMPD charcoal-based 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 results in the data becoming accessible to the end user through the WDS-Paleo Dataset Search: https://www.ncei.noaa.gov/paleo-search/?dataTypeId=12, as well as other Data Access tools located at: https://www.ncei.noaa.gov/products/paleoclimatology/fire-history

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