

Example Data Management Plan: NSF General

Product of a DataONE Best Practices Workshop, based on the work of C.D. Keeling and Colleagues, Scripps Institution of Oceanography.

Background Information (not part of a DMP)

Title:

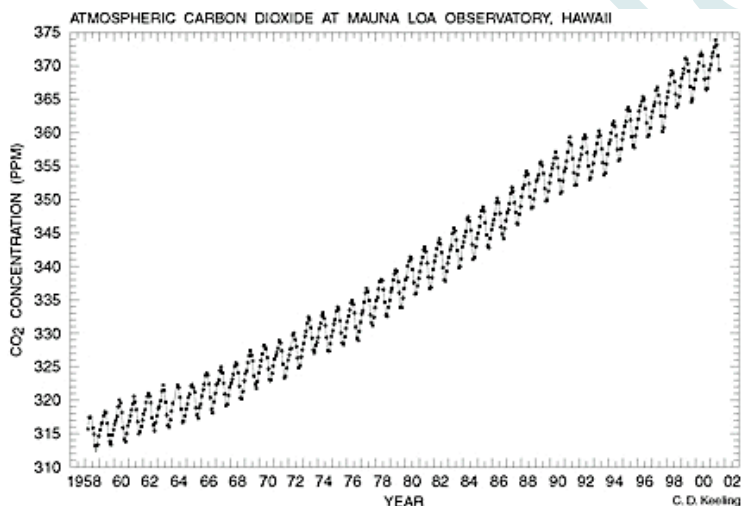
Atmospheric CO₂ Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013.

Purpose of project:

The purpose of this proposed project is to study the controls on the concentration of atmospheric CO₂ using high precision and accuracy measurements at a remote island observatory.

We propose to measure the concentrations of CO₂ in the atmosphere at the Mauna Loa Observatory, Hawaii. The methodology for sample collection and analysis during this project will generate highly accurate and precise data that can be seamlessly added to the existing Mauna Loa CO₂ record (1958 – 2010) [1,2]*. A major theme for this project is to identify and minimize systematic measurement errors through rigorous sampling and calibration procedures.

We have chosen to use an iconic data product—the Keeling Mauna Loa CO₂ record—for this example Data Management Plan. All environmental scientists are familiar with this data record. It is posted in the atrium of the U.S. National Academy of Sciences, next to the DNA Double helix. We are writing this Data Management Plan as if it were to be included in an NSF proposal in 2011.



* Citations point to references located within the proposal body.

Data Management Plan

Atmospheric CO₂ Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013.

1. Types of Data Produced

Air samples at Mauna Loa Observatory will be collected continuously from air intakes located at five towers – a central tower and four towers located at compass quadrants. Raw data files will contain continuously measured CO₂ concentrations, calibration standards, references standards, daily check standards, and blanks. The sample lines located at compass quadrants were used to examine the influence of source effects associated with wind directions [3,4]. In addition to the CO₂ data, we will record weather data (wind speed and direction, temperature, humidity, precipitation, and cloud cover). Site conditions at Mauna Loa Observatory will also be noted and retained.

The final data product will consist of 5-minute, 15-minute, hourly, daily, and monthly average atmospheric concentration of CO₂, in mole fraction in water-vapor-free air measured at the Mauna Loa Observatory, Hawaii. Data are reported as a dry mole fraction defined as the number of molecules of CO₂ divided by the number of molecules of dry air multiplied by one million (ppm).

The final data product has been thoroughly documented in the open literature [2] and in Scripps Institution of Oceanography Internal Reports [1].

The data generated (raw CO₂ measurements, meteorological data, calibration and reference standards) will be placed in comma-separated-values in plain ASCII format, which are readable over long time periods. The final data file will contain dates for each observation (time, day, month and year) and the average CO₂ concentration. The final data product distributed to most users will occupy less than 500 KB; raw and ancillary data, which will be distributed on request comprise less than 10 MB.

2. Data and Metadata Standards

Metadata will be comprised of two formats—contextual information about the data in a text based document and ISO 19115 standard metadata in an xml file. These two formats for metadata were chosen to provide a full explanation of the data (text format) and to ensure compatibility with international standards (xml format). The standard XML file will be more complete; the document file will be a human-readable summary of the XML file.

3. Policies for Access and Sharing

The final data product will be released to the public as soon as the recalibration of standard gasses has been completed and the data have been prepared, typically within six months of collection. There is no period of exclusive use by the data collectors. Users can access documentation and final monthly CO₂ data files via the Scripps CO₂ Program website (<http://scrippsco2.ucsd.edu>). The data will be made available via ftp download from the Scripps Institution of Oceanography Computer Center. Raw data (continuous concentration measurements, weather data, etc.) will be maintained on an internally accessible server and made available on request at no charge to the user.

4. Policies for Re-use, Distribution

Access to databases and associated software tools generated under the project will be available for educational, research and non-profit purposes. Such access will be provided using web-based applications, as appropriate.

Materials generated under the project will be disseminated in accordance with University/Participating institutional and NSF policies. Depending on such policies, materials may be transferred to others under the terms of a material transfer agreement.

Publication of data shall occur during the project, if appropriate, or at the end of the project, consistent with normal scientific practices. Research data which documents, supports and validates research findings will be made available after the main findings from the final research data set have been accepted for publication.

5. Plans for Archiving and Preservation

Short Term:

The data product will be updated monthly reflecting updates to the record, revisions due to recalibration of standard gases, and identification and flagging of any errors. The date of the update will be included in the data file and will be part of the data file name. Versions of the data product that have been revised due to errors / updates (other than new data) will be retained in an archive system. A revision history document will describe the revisions made.

Daily and monthly backups of the data files will be retained at the Keeling Group Lab (<http://scrippsco2.ucsd.edu> , accessed 05/2011), at the Scripps Institution of Oceanography Computer Center, and at the Woods Hole Oceanographic Institution's Computer Center.

Long Term:

Our intent is that the long-term high quality final data product generated by this project will be available for use by the research and policy communities in perpetuity. The raw supporting data will be available in perpetuity as well, for use by researchers to confirm the quality of the Mauna Loa Record. The investigators have made arrangements for long-term stewardship and curation at the Carbon Dioxide Information and Analysis Center (CDIAC), Oak Ridge National Laboratory (see letter of support). The standardized metadata record for the Mauna Loa CO₂ data will be added to the metadata record database at CDIAC, so that interested users can discover the Mauna Loa CO₂ record along with other related Earth science data. CDIAC has a standardized data product citation [5] including DOI, that indicates the version of the Mauna Loa Data Product and how to obtain a copy of that product.