Example Data Management Plan, based on the work of C.D. Keeling and Colleagues, Scripps Institution of Oceanography

Background (not part of the DMP)

A hypothetical proposal to NSF:

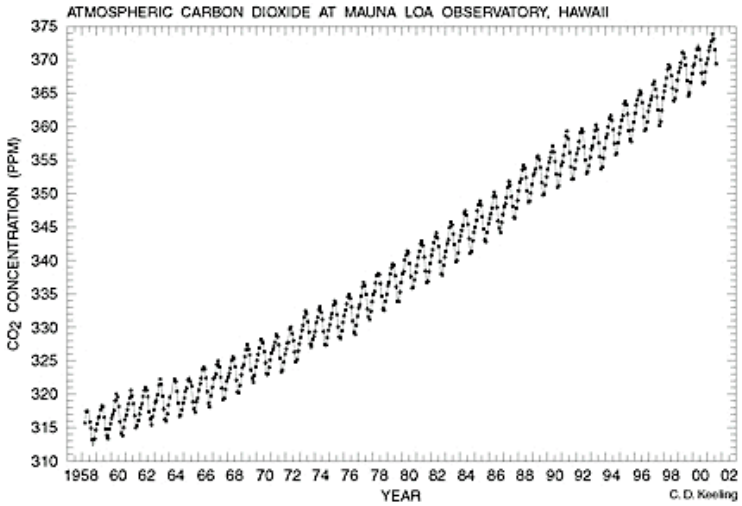
Atmospheric CO_2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013.

The purpose of this hypothetical project is to study the controls on the concentration of atmospheric CO_2 using high precision and accuracy measurements at a remote island observatory.

For this hypothetical project, we propose to measure the concentrations of CO_2 in the atmosphere at the Mauna Loa Observatory, Hawaii. The methodology for sample collection and analysis during this proposed project will generate highly accurate and precise data that can be seamlessly added to the existing Mauna Loa CO_2 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 CO_2 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.

This example data management plan illustrates the key elements required for a two-page plan for a NASA or NSF proposal.



Data Management Plan

1. Information about the data

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 CO_2 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 CO_2 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 CO_2, 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 CO_2 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 CO_2 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 CO_2 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, will occupy less than 10 MB.

2. Metadata content and format

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. Short-term storage and data management

The data product will be updated monthly due to updates to the record, revisions due to recalibration of standard gases, and due to 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.

4. Policies for access and sharing

The final data product will be release 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 CO_2 data files via the Scripps CO_2 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.

5. Long-term storage and data management (archiving)

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 CO_2 data will be added to the metadata record database at CDIAC, so that interested users can discover the Mauna Loa CO_2 record along with other related Earth science data. CDIAC has a standardize 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.

References: (would be located with other References in Proposal body)

[1] Keeling, CD, SC Piper, RB Bacastow, M Wahlen, TP Whorf, M Heimann, and HA Meijer, **2001**. Exchanges of atmospheric CO_2 and $^{13}CO_2$ with the terrestrial biosphere and oceans from 1978 to 2000. I. Global aspects, SIO Reference Series, No. 01-06, Scripps Institution of Oceanography, San Diego CA, 88 pages.

[2] Keeling, CD, SC Piper, RB Bacastow, M Wahlen, TP Whorf, M Heimann, and HA Meijer, **2005.** Atmospheric CO_2 and $^{13}CO_2$ exchange with the terrestrial biosphere and oceans from 1978 to 2000: observations and carbon cycle implications, In *A History of Atmospheric CO_2 and its effects on Plants, Animals, and Ecosystems*. Edited by JR Ehleringer, TE Cerling, and MD Dearing, Springer Verlag, New York, pages 83-113.

[3] Bacastow, RB, Keeling, CD and Whorf, TP, **1985**. Seasonal amplitude increase in atmospheric $C0_2$ concentration at Mauna Loa , Hawaii, 1959-1982. J. Geophys. Res. 90, 10529-10540.

[4] Sundquist, ET and RF Keeling, **2009**. The Mauna Loa Carbon Dioxide Record: Lessons for Long-Earth Observations. In *Carbon Sequestration and Its Role in the Global Carbon Cycle*. AGU Monograph Series 183. Edited by BJ McPherson and ET Sundquist. pages 27-35. doi: 10.1029/2008GM000713.

[5] Keeling, RF, SC Piper, AF Bollenbacher and JS Walker, **2009**. Atmospheric CO_2 records from sites in the SIO air sampling network. In *Trends: A Compendium of Data on Global Change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge,TN, doi: 10.3334/CDIAC/atg.035

N.B. A later version of this document tailored specifically to data management plans for NSF proposals is available from DataONE