Data Management Plan
Sample Unidata NSF Project

About the Project Data

This is an example of a Data Management Plan based on the NSF Division of Atmospheric and Geospace Sciences (AGS) template. Here we provide a sample plan for an imaginary project that collects data from a network of data loggers and creates a collection of time-series plots for analysis.

This document is not meant to be used as-is, but rather to provide you with ideas about how to structure a Data Management Plan for an NSF proposal, and how you might use Unidata technologies to implement the plan.

Note: Data Management Plans for NSF proposals must be two pages or less when submitted.

The "About the Project" section provides a brief description (what, where, when) of the data that will be produced by the project.

This imaginary project will collect new observational data about air temperature, humidity, wind speed, barometric pressure, and time from 20 locations in Boulder County, Colorado over the space of 16 weeks in March, April, May, and June of 2015. The data will be collected using Bellcamp RC008 dataloggers.

Approximate Data Volume

The dataloggers will record values at 30-minute intervals, 24-hours per day. Each datalogger will generate approximately 1.5 Mbytes of data per day in ASCII comma-separated values (CSV) format. The resulting volume of collected data (from all 20 dataloggers) will be approximately three to four Gbytes over the duration of the project.

After collection, the data will be transformed from ASCII format to netCDF-4 format using Unidata's Rosetta data translation tool. We expect the resulting netCDF data set to be approximately one Gbyte.

Metadata

In addition to the quantities being measured, the Bellcamp RC008 dataloggers will record sensor calibration information for each attached sensor device. Sensor calibration information will be recorded as metadata in the resulting netCDF files.

During transformation from ASCII to netCDF with the Rosetta data translation tool, researchers will add the following metadata to each record: Sensor location, units for each physical quantity recorded, time offset from UTC, and record series number.

The following project metadata will be added to the netCDF collection itself: Project title, short project summary, contact information for project lead, names of project contributors, and NSF grant information.

Data Availability

ASCII data from the project dataloggers will be collected daily by physically connecting laptop computers to the dataloggers. Daily data from the laptop computers will be uploaded to the project's RAMADDA server at the end of each day's collection. ASCII files will be transformed to netCDF using Rosetta on a weekly schedule.

During the data collection (16 weeks) and analysis (8 weeks) phases of the project, data will only be available to the project team.

After the analysis phase of the project, all netCDF data and analysis figures will be made available to other researchers/reviewers for viewing or download by request via password-protected login to the project's RAMADDA server.

After publication of project results, all netCDF data and analysis figures will be made available to the public via anonymous access to the project's RAMADDA server.

Use Policies

After publication of project results, the project data will be freely available for re-use, distribution, or production of derivative works as long as attribution of the source of the project data, including names of project researches and information on any grants under which the project was undertaken, are included in the new product. Specific attribution guidelines will be included on the project's RAMADDA site.

Data Archiving

Project data will be kept available on the project RAMADDA server for a minimum of two years after publication of the project results.

Two years after publication of the project results, the data archive will be submitted to the National Climatic Data Center (NCDC) for long-term archiving. The data may still be available on the project RAMADDA server after submission to NCDC, but this is not guaranteed past the time when data is accepted into the archive. If data is no longer available on the project RAMADDA server, reasonable attempts will be made to provide pointers from there to the data archive.

Note: Assigning a Digital Object Identifier (DOI) to the data set would make it easier for future researchers to locate the data once it is archived.